TITLE: Technologies For Large-Scale Physical Mapping Of Human Chromosomes

AUTHOR(S): Tony J. Beugelsdijk

SUBMITTED TO: International Conference on Automation,
Robotics, and Artificial Intelligence
February 7-10, 1995
Montreaux, Switzerland

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes.

The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

"Technologies for Large-Scale Physical Mapping of Human Chromosomes"

Tony J. Beugelsdijk, Ph.D., M.B.A.

**Engineering Sciences & Applications Division
and
Center for Human Genome Studies
Los Alamos National Laboratory
Los Alamos, NM 87545**

Introduction

Since its inception 6 years ago, the Human Genome Project has made rapid progress towards its ultimate goal of developing the complete sequence of all human chromosomes. This progress has been made possible through the development of automated devices by laboratories throughout the world that aid the molecular biologist in various phases of the project. The initial phase involves the generation of physical and genetic maps of each chromosome. This task is nearing completion at a low resolution level with several instances of very high detailed maps being developed for isolated chromosomes.

In support of the initial mapping thrust of this program, the robotics and automation effort at Los Alamos National Laboratory has developed DNA gridding technologies along with associated database and user interface systems. This paper will discuss these systems in detail and focus on the formalism developed for subsystems which allow for facile system integration.

The DNA Filter Array Problem Description

Generation of filter arrays is shown schematically in Figure 1. Also shown is the tracking scheme employed for individual colonies. The physical task involves making an imprint of the contents of a 96 well microtiter plate onto a nylon filter. The tool used has 96 pins which are lowered by a positioning system into the source plate and then transferred to the filter. The spot-to-spot spacing on the filter is the same as the well-to-well spacing in the source microtiter plate. This distance is typically 9 mm. The process is repeated using a clean or sterilized tool in making a subsequent impression from a second plate.

Figure 1. Array Construction and Colony Tracking Scheme.

However, rather than moving the tool to a clean part of the filter, it is shifted slightly over the first impression. This process can then be repeated many times with successive plates until the filter is filled. Filling a filter can proceed to quite high interleave densities. Indeed 36 to 1 has been reported. For our purposes, we use a 16 to 1 interleave density. Since these spots are bacterial cultures which are then incubated overnight, they grow into each other precluding their initial positioning at higher interleave densities.

Also depicted in Figure 1 is the colony tracking scheme shown, by way of example, using a 4 to 1 interleave density. Each source plate has a unique barcode label. As each plate is presented to the system, its label is scanned. Individual colonies in each plate are assigned a number based on the plate number with a row and column suffix designating their position. In the center of Figure 1 is a filter constructed using 4 plates. Reading across the first 2 spots are numbers 12345A01 and 23456A01, respectively. Similarly, the first 2 colonies reading down in the first column are 12345A01 and 34567A01, respectively. It is important to note that no two adjacent spots on the completed filter originate from the same source plate. This becomes increasingly important when imprinting at higher interleave densities.

Since it is the use of these filter arrays in subsequent experiments that is of real interest, the task of generation of DNA filter arrays requires careful definition prior to hardware and software design.

Shown in Figure 2 is a schematic depicting the various elements of filter array construction.

Figure 2. Filter Array Problem Definition.

At the center is the physical device that constructs the filter array. When we initiated this development in 1988, no such systems were commercially available. This was also the case in 1991 when we built the second generation system. Today, there are several systems that are either custom built or commercially available that will construct filter arrays. The distinction between systems is in the amount of operator attention required and in the subsequent experiments involving the filters.

At the next level of problem definition is the database of colonies, positions, identities, etc. With a totally automated system, it is possible to generate many hundreds of filter arrays each day each with 1536 (16 to 1 interleave density) spots of DNA, the data handling logistics become very important.

The next level is the user interface. Ideally, it is graphical and intuitive. The overall system derives its utility from being used as a tool to achieve an end. As such it should not become the centerpiece of the operator's attention and be easy to use.

Finally, a data interpretation tool is necessary to deal with the images generated from the filter arrays. The filters are subsequently subjected to hybridizations with various probes. These experiments result in an image of selected spots on an x-ray film. Therefore image analysis tools working in conjunction with database tools allow for the rapid identification of any spots and its tracking to a particular colony in a source plate.

The various components depicted in Figure 2 are all required for an integrated filter arraying capability and have been incorporated into our designs for both the 1st and 2nd generation gridding systems. These systems will be described below.

First Generation DNA Filter Array System

The first generation DNA filter array system¹ developed is shown in Figure 3. The system consists of 2 microtiter plate stackers - one for presenting plates to the robot and the other for return of the plates already arrayed, a barcode reader, 2 filter stands each capable of 4 filter arrays, tool washing baths, and a small table top gantry robot. The robot is manufactured Nutec and is supplied with a Motion Systems controller. The host computer is an IBM PC.

Figure 3. First Generation DNA Filter Array System.

Up to 16 plates can be placed in the presentation stacker in any order. As each is presented, its barcode label is scanned. The lid is then removed and a clean tool is dipped into the wells. The robot is directed by the database² on the placement position on the filter. In this way, all filter arrays are identical regardless of the order of the source plates. Moreover, if the system fails in any way, an operator can intervene, place the source plate back on top of the stack and continue. Most other systems that experience failures cannot recover and have to be reconfigured and reinitialized. For these systems, all work up to the point of failure is lost.

In this first generation system, arrays of 96, 384, 864, or 1536 clones can be constructed representing, respectively, no, 4 to 1, 9 to 1, and 16 to 1 interleave densities. Eight filter arrays can be constructed in a 2 hour unattended run. The database which records individual colony locations is constructed simultaneously.

This system has been run for several years in a production mode. IN an average day, 24 filter arrays can be produced. Well over a thousand arrays have been generated for internal use as well as for LANL collaborators.

Limiting this throughput is the fact that the robot is directed to clean the tool after every impression even if it goes back to the same source plate. The robot also has to move the tool between successive cleaning baths which is time-consuming. These limitations and others were addressed in the second generation instrument.

Second generation DNA Filter Array System

Since tool cleaning was time-consuming and rate limiting in the initial system, a tool cleaning station was designed and incorporated. A schematic of this station is shown in Figure 4.

Figure 4. Automatic tool cleaning station.

In operation the robot places a used tool in the left position and detaches the tool. It then picks up a clean tool from the position at the far right. The tool cleaning station then proceeds to raise an ultrasonic bath under the dirty tool. After a prescribed cleaning period, the bath is lowered and a pneumatic actuator pulls the tool one position to the right where a warm air current flows over the tool to dry the tips. The next cleaning cycle forces this tool to an intermediate cooling position, and the final cleaning cycle places this tool in a cool air stream ready for pickup by the robot.

Another subsystem that was redesigned for the second generation instrument was the plate stacker. The initial ones were purchased commercially and extensively modified to be reliable in this application. Even though, occasional failures occurred necessitating human intervention.. These deficiencies were corrected in a new plate stacker shown in Figure 5. This stacker performs the same functions as the initial ones, but much more reliably. A barcode reader is internal to the base housing. This stacker has a capacity of 32 microtiter plates. With extensions on the columns, about 50 plates can be accommodated.

The individual tools were also redesigned to make for easier changing of the tips. Different tip geometries are best suited for cosmids, yeast artificial chromosomes or YACs, phage, and naked DNA. The tools can be attached and detached by the robot through the use of the Applied Robotics Xchange™ tool changing system.

Finally, a different robot was used with a larger work envelope. The robot chosen was the Adept 604S, which has a 3 second cycle time³ and has a repeatability of 0.002 inches. This robot is used extensively in the semi-conductor industry and is known for its speed and reliability. The work envelop can accommodate up to 5 filter stands, with each filter capable of 4 filter arrays. An isometric view of this system is show in Figure 6 and a recent picture of the system is shown in Figure 7.

The control structure was also changed significantly in this system from that of the first. This will be described below in the discussion on the integration formalism employed.

Figure 5. Microtiter Plate Stacker.

Figure 6. Isometric View of Second Generation Filter Array System.

Figure 7. Adept 604S Based Filter Array System.

Data Interpretation Tools

The filter arrays are processed further by fixation of the DNA onto the filter and denaturation of the double strands. They are then washed with a solution containing a radio-labelled oligonucleotide probe. Wherever this probe finds a complementary strand of DNA on the filter it hydrogen bonds or hybridizes to that strand. In the process, a radio-labelled spot is created on the filter. Excess probe solution is washed from the filter which is then dried and exposed for a length of time to a piece of x-ray film. A typical image which results is shown in Figure 8.

Figure 8. Autoradiograph of a Filter Array Hybridization.

At this point, a biologist "reads" the image by scoring the spots as negative or various degrees of positive. The positive spots correspond to unique colonies in each of the source microtiter plates that were used to generate the filter array. Whenever, 2 colonies are positive, they share a piece of DNA in common. This information forms the basis for constructing an overlapping physical map of the chromosome.

To assist the biologist in the data interpretation task, a software tool called GridHyb was developed. This tool was developed in 2 stages. The initial tool was for computer assisted data entry for each hybridization experiment with the biologist making the scoring assignments. A second phase incorporates automatic scoring. Both modes allow for the biologist to review and edit the scores as necessary. The GridHyb software combines the database generated during the construction of the filter array with a user interface and image analysis tools into a single package.

The initial version of the GridHyb tool has the following functions: 1) it automates data entry for each experiment, 2) it retrieves data from past experiments for review, 3) it displays digitized images of auto radiographs, and 4) it provides computer assisted scoring of the images. The software is written in C using Motif™ for Unix workstations which run X-Windows™.

A typical user interface screen is shown in Figure 9. Depicted here are the 4 filter arrays. Each filter array is composed of sources material from 16 microtiter plates numbered here from 1 to 16, 17 to 32, etc. Buttons across the top of the screen allow access to the Human Genome Information Resource (HGIR). It also gives information about the probe session which is defined as a combination of filter, plates, probe identity, etc. for each hybridization experiment. Since it is possible to reuse the same filter for several hybridization experiments, it is necessary to assign each one a unique probe session identifier.

The screenshot displays the GridHyb User Interface. At the top, there are five buttons: "Log In To HGIR", "ProbeSession Infor", "Grid Information", "Probe Information", and "Log Out". Below these is a status bar showing "Probe Session: IRSyGHybGriddedRearray[1.0]" and "Grid: Probe:". Underneath the status bar are four buttons: "Select Probe Session", "New Probe Session", "Next Session", and "Program Help", followed by a "Quit" button. The main area contains four filter arrays, each represented as a 4x4 grid of numbered squares. The arrays are labeled 1.0, 2.0, 4.0, and 3.3. Below each array, there is a label "HGIR = USER =". At the bottom, there is a text box that says "Selecting a Filter Toggle Button Enables the Editing Mode and Pops Up the Grid Window". Below this text box are four toggle buttons labeled 1.0, 2.0, 3.3, and 4.0.

Filter Array	Grid 1 (1-16)	Grid 2 (17-32)	Grid 3 (33-48)	Grid 4 (49-64)
1.0	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32	33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48	49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64
2.0	HGIR = 11 USER = 4	HGIR = 0 USER = 26	HGIR = 5 USER = 0	HGIR = 9 USER = 4

Selecting a Filter Toggle Button Enables the Editing Mode and Pops Up the Grid Window

1.0 2.0 3.3 4.0

Figure 9. GridHyb User Interface Screen.

Across the bottom of the screen are a series of toggle buttons which allow the user to open any of the filters for editing. If button 2 is selected the screen shown in Figure 9 is presented. This cartoon shows each spot in a 16 to 1 interleave density filter array. In use, the operator places the actual autoradiograph for this experiment on a light box or superimposes this cartoon over the actual digitized image of the autoradiograph. At the right of the window are symbols representing the various scores that can be assigned to varying levels of darkness. The boundaries of the bins into which the spots are sorted can be set by the operator. As each clone is scored by the operator, its identity appears in the window labelled "Unsaved Edited Clones." Also in the upper window more detailed clone information is available.

The second phase of the GridHyb tool development is the automated scoring of the filter array images. This tool provides an extension to the existing tool and features automatic aligning of the cartoon to the image, automatic ranking of all the spots, customization of the bin boundaries, and an editing mode.

Figure 10. GridHyb User Interface - Filter Array Cartoon.

An unprocessed image such as the one that is shown in Figure 7 is given as input to this tool. Since the filters can stretch with subsequent probings, this tool reconstructs the rectilinear array. Extraneous spots often show up in the image and are discarded. Automatic aligning of the image occurs in stages with successive thresholding on the image. The darkest spots show up on the first level thresholded image. Successive thresholded images are processed until unambiguous array locations are assigned. Typically, this takes about 2 passes through the image. At this point the tool makes suggested scoring assignments which can be edited or taken as assigned.

Integration Formalism 4,5

Each automation effort seems to be a grounds up development of hardware and software. To avoid re-developing key components of any system, we have developed the Standard Laboratory Module (SLM) formalism for interconnecting subsystems in an automated instrument. By constructing these subsystems to a pre-defined interface standard, interconnection into a working system is greatly simplified.

At the module or SLM level a combination of unit operations are physically linked together in a device such that this device performs a complete subtask within a protocol. In the instance of DNA filter array construction, such a subtask might be presenting a plate, or cleaning a tool, or making a filter impression. These tasks are then incorporated into a hardware and/or software module, i.e., an SLM, which is then integrated with other SLMs to make a complete working system.

At the heart of the SLM is the interface definition. Each SLM is a finite state machine. The finite state diagram is shown in Figure 10.

Figure 10. SLM State Diagram.

An SLM is always in one of these states. What distinguishes one SLM from the next is the sequence that is executed up entering the "normal operation" state. With these states defined, it becomes straightforward to integrate many SLMs into a working system. The sequencing of individual SLMs is accomplished with a Task Sequence Controller or TSC. A user interface has been generated which allows for interaction with the TSC as well as a bench manager and an external database. These system elements are shown in Figure 11.

Each SLM has its embedded controller which is well-behaved according to the state diagram and responds predictably to commands from the TSC. This architecture is generic in that it applies to all laboratories where samples are brought in and information is generated from them. We have been successfully applying this technology to

environmental laboratories handling a wide variety of samples from soils, to sludges to waters and have begun using this approach in genetic and clinical applications.

Figure 12. SLM Integration Architecture.

References

1. Patricia A. Medvick, Robert M. Hollen, and Randy S. Roberts, "Development of an Automated Workcell for DNA Hybridization Array Construction," *J. Laboratory Automation and Robotics*, 3, 169-173 (1991).
2. P. A. Medvick, R. M. Hollen, R. S. Roberts, D. Trimmer, and T. J. Beugelsdijk, "Automated DNA Hybridization Array Construction and Database Design for Robotics Control and For Source Determination of Hybridization Responses," *International Journal of Genome Research*, 1(1), 17-23(1992).
3. A standard cycle time is the time required to move a 1 pound item up 1 inch, over 12 inches, down 1 inch and return to the starting point retracing the same path.
4. A complete description of this program is found in the *Journal of Laboratory Automation and Robotics*, Vol. 6, No. 2, 1994.
5. T. J. Beugelsdijk, R. M. Hollen, and T. H. Errkila, "The Standard Laboratory Module: An Integrated Approach to Standardization in the Analytical Laboratory," *Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management*, 21,207-214(1993).

Acknowledgements

The author would like to acknowledge the contributions of Drs. Patricia A. Medvick, Randy S. Roberts, technician David C. Bruce, and graduate students Catherine C. Beauheim, Bobi K. Den Hartog, and Frank S. Smutniak.