DNA SEQUENCING

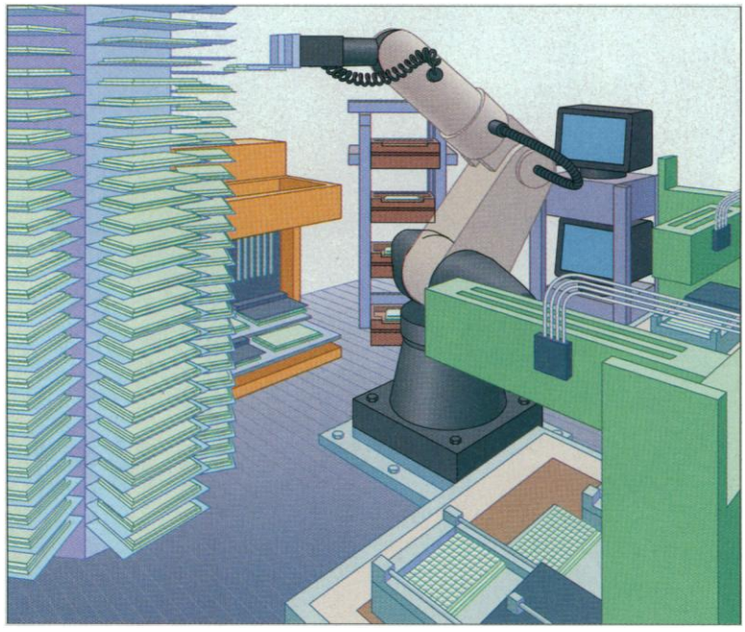
A Magnetic Attraction to High-Throughput Genomics

Trevor L. Hawkins, Kevin J. McKernan, Laurent B. Jacotot, Jason B. MacKenzie, Paul M. Richardson, Eric S. Lander

In the two decades since the invention of DNA sequencing, only a tiny fraction of the human genome has been decoded—about 1% of the 3000 mega–base pair (Mbp) total. It is therefore somewhat daunting to consider that the Human Genome Project has committed to completing the remaining 99% of the sequence by the year 2005 (*1*). This will require an unprecedented increase in sequencing capability, with a handful of genome centers needing to produce a combined throughput of 500 Mbp per year. In practical terms, this translates to processing 50,000 samples per day through the laboratory steps of DNA purification and sequencing, as well as tracking and analyzing the resulting 15 gigabytes of data generated each day. With this in mind, we designed an automated sequencing system to meet the laboratory throughput needs of the Human Genome Project.

TECH VIEW

We began by considering the key performance requirements: (i) Mechanically, the system must manipulate thousands of microliter-volume samples through different and changing biochemical procedures. Although highly integrated chip-based systems for sample processing are under development, the most reliable way to accomplish the goal with current technology is to emulate the human manipulations performed in the laboratory. With an articulated arm at the center of a set of modules for liquid handling, thermocycling, shaking, and storage, all coordinated by scheduling software, we were able to create a generic automation platform. (ii) Biochemically, the system should avoid the traditional, but hard-to-automate, methods of centrifugation and solvent extractions used in DNA isolation. To overcome these limitations, we developed a procedure called solid-phase reversible immobilization (SPRI) (*2*). Under certain conditions, DNA can be tightly bound to the surface of carboxyl-coated magnetic particles, extensively washed, and subsequently released back into solution. (iii) Procedurally, the system must be highly flexible to incorporate both minor optimization and major redesign of biochemical steps. This demands a sophisticated scheduling and control system communicating with a sample-tracking database.



ANTHONY DAVIS

Sequatron sequencing. In the center is the CRS A465 robotic arm, shown transferring a microtiter plate from the carousel (left) to one of the two Packard 104 XYZ robots (green) that set up the sequencing reactions. The Tecan XYZ robot (orange, to the right of carousel) is used to pool completed sequencing reactions, and the Techne and MJ Research thermal cyclers (brown) are used for performing the sequencing reactions.

These considerations led us to design the Sequatron, an automated, adaptable system for high-throughput genomics. The term Sequatron actually describes a generic platform with an articulated robotic arm, centralized control, and scheduling software. An application can be built around this flexible base and customized by placing different modules onto the workspace. The Sequatron systems were initially designed for DNA sequencing, but also provide a general solution to tasks in molecular biology.

The basic Sequatron platform consists of a table 1.5 by 1.5 m with a CRS A465 articulated arm (CRS Robotics, Burlington, Canada) with customized fingers attached to the wrist (see figure). A central personal computer (PC) directly controls the arm and sends instructions to the other devices (or to their computer controllers) via serial connections (see diagram). The PC also schedules events, such as moving plates to and from liquid handling systems or thermal cyclers; scheduling is made possible with a convenient programming language developed by CRS Robotics. This allows methods to be modified or optimized without significant changes to the control software. Finally, the PC communicates to process-tracking databases via an Ethernet link; information on each run is downloaded from the database and the results are uploaded back to the database.

At present, we have three production Sequatron systems to perform (i) DNA purification from M13 phage and polymerase chain reaction (PCR) products, (ii) DNA sequencing reactions, using dye-primer and dye-terminator chemistries, and (iii) finishing, which involves performing custom PCR amplifications and sequencing reactions on selected templates in order to close gaps in a partially completed genomic sequence. (The third system has been created in prototype, but is still under active development.)

The purification Sequatron isolates DNA from M13 and PCR products with SPRI. The challenge here is to achieve reproducibility and economy. The magnetic bead approach has proven to be highly robust and cost effective (less than 10 cents per template) in all our applications and has become the generic method for DNA isolation and manipulation for the Sequatron systems. The purification Sequatron has a throughput of approximately 16,000 templates per day with unattended operation.

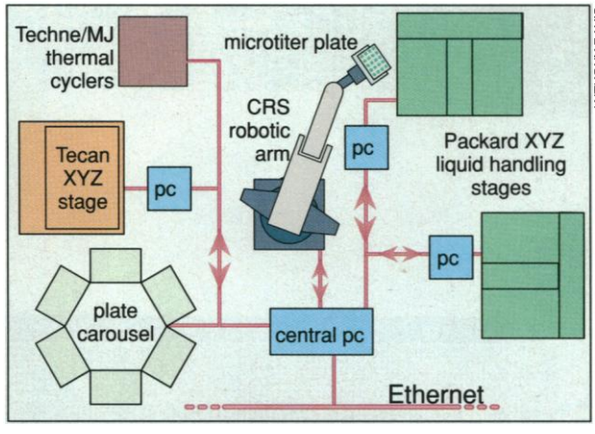
The sequencing Sequatron performs DNA sequencing reactions with M13 DNA purified by the previous system. Automation of se-

The authors are at the Whitehead Institute, Massachusetts Institute of Technology Center for Genome Research, Cambridge, MA 02141, USA. E-mail: tlh@genome.wi.mit.edu

quencing reactions requires unattended operation of liquid handling in 384-well plates, thermocycling of reactions, and pooling back into 96-well plates. The Sequatron uses three types of modules to set up the sequencing reactions. The first is a Packard 104 XYZ liquid-handling robot (Packard Instrument, Meriden, CT), modified by replacing one of its four single-pipette tips with a 12-channel pipette; the 12 new tips are linked to two six-channel dispensers (Cavro, Sunnyvale, CA) controlled by the Packard robot control unit via a serial link. The result is a highly flexible workstation capable of rapid dispensing to and from microtiter plates, as well as from other reagent sources. The second type of module is a liquid handler (Tecan US, Research Triangle Park, NC), also modified to use a single 12-channel pipette. The third type of module is a custom thermal-cycling device (Techne, Princeton, NJ, and MJ Research, Watertown, MA), having 384-well Peltier blocks with heated lids controlled through serial connections. The advantage of a 384-well thermocycler is that DNA templates from a 96-well plate can be conveniently split into the four separate sequencing reactions (for nucleotides A, T, C, and G) required for dye-primer sequencing, which simplifies sample tracking and reduces costs. The sequencing Sequatron has been in operation for the past 3 months, producing about 2000 sequencing reactions in a single 3-hour run per day. Running for an entire workday increases the throughput to more than 16,000 samples.

The finishing Sequatron combines all the tasks in the previous systems, but has more modules and lower throughput because of the need to perform a range of different sequencing chemistries on a small percentage of templates. Finishing involves filling gaps in genomic sequence, which requires performing customized steps on strategically located clones; the steps include reverse reads from the opposite end of clones, primer walks to extend sequences, and dye-terminator chemistries to resolve sequence ambiguities. The system selects clones on the basis of instructions from TaskMaster, our laboratory information management system, and schedules a customized Sequatron run. A typical run may involve selecting about 500 clones from among roughly 10,000 clones, setting up PCR reactions where necessary, and performing various sequencing chemistries.

The Sequatrons have advantages in terms of cost, quality control, and work-flow management. Two Sequatrons produce 16,000 samples per day with minimal human intervention; traditional manual methods would require significantly more employees. The second benefit of these systems is their ability to monitor quality from run to run and to optimize protocols in a controlled manner without human inconsistencies. Finally, the system allows an organized workflow of sample batches moving from Sequatron to Sequatron, with results tracked online. As with any automation, one disadvantage of the Sequatron is the time and investment needed to build and maintain the system, especially given the rapidly evolving methods used in DNA sequencing. Despite our efforts to make the system flexible, it is locked into certain processes or methods. For example, if we were to move away from the use of magnetic beads, we would have to make significant mechanical changes to the systems. However, the Sequatron's ability to move microtiter plates or any similar objects should enable the system to be used for a multitude of tasks as new methods are developed.



ANTHONY DAVIS

Components in the sequencing Sequatron. Each module is controlled by a central computer that is tied by Ethernet to a process-control database.

Few comparable integrated robotic systems exist for high-throughput sequencing. Low-throughput systems for DNA purification and sequencing have been developed by manufacturers of XYZ robotic units, including Applied BioSystems (Foster City, CA), Tecan, and Packard; these systems are more suited to human-assisted mechanization than complete automation. The Sequatron more closely resembles the sort of systems used by pharmaceutical companies for high-throughput drug screening. For example, Beckman Instruments (Fullerton, CA) recently introduced an integrated system, combining Beckman XYZ robots, SAGIAN (Indianapolis, IN) robotic arms, and detection units from Molecular Devices (Sunnyvale, CA). Because the genome-sequencing market is smaller and more specific than even the drug-screening market, commercial high-throughput integrated systems have not been available. Accordingly, the Sequatron was designed and has continued to evolve to meet the needs of the Human Genome Project. As the project advances, it will be important to transfer such technologies from larger genome centers into the hands of researchers in both academic and industrial environments.

Future challenges include reducing the volume of reagents (a major cost in DNA sequencing) and directly coupling to a sequence detector (to avoid manual loading of sequencing gels, one of the most labor-intensive steps in the process). We have achieved a 10-fold reduction in volume by using a liquid-handling system with piezoelectric dispensers (BioChip Processor, Packard Instrument), but the use of such small volumes introduces the problem of evaporation and may prompt the need for integrated chip-based systems for liquid handling. We have also experimented with ways to couple sample preparation to sequence detectors. The Sequatron naturally interfaces with capillary-based or mass spectrometry–based sequencers, which use microtiter plates. Alternatively, the Sequatron could be used with existing slab gel sequencers (such as the Applied BioSystems model 377) by mounting a system on an autonomously guided vehicle able to move and position the arm in front of a sequencer. More generally, the Sequatron system provides a flexible platform for other automation needs, both within and beyond the Human Genome Project. Sequatron prototypes are in development for construction of subclone libraries, pooling of bacterial artificial chromosome (BAC) and other libraries, and characterization of large-insert clones to be sequenced. Future systems could also be directed toward such functional genomic tasks as genotyping, setup of high-density arrays, and high-throughput screening of pharmaceutical compounds. At some point in the future, smaller Sequatron systems performing molecular biology tasks could well be in typical lab environments.

References and Notes

1. F. Collins and D. Galas, *Science* **262**, 43 (1993).
2. T. L. Hawkins, T. O'Connor, A. Roy, C. Santillan, *Nucleic Acids Res.* **22**, 4543 (1994).

3. We thank the members of the sequencing group, especially A. Christoffersen, M. P. Reeve, and W. Lee for advice about the design of these systems. We also thank R. Papen (Packard Instrument) for help with the liquid handling systems. Supported by grants to T.L.H. from the Department of Energy (DE-FG02-95ER62099) and to E.S.L. from NIH (HG00098).

TechWire Forum:
www.sciencemag.org/dmail.cgi?53201

A Cell Growth Switch

SIGHTINGS

Researchers have long sought biochemical switches that turn cell growth on and off. Control over these functions would have clinical uses as well as implications for basic cell biology and biotechnology. In a recent issue of *Proceedings of the National Academy of Sciences* (*1*) Blau *et al.* have reported what may turn out to be a powerful tool for controlling cell growth in vitro and in vivo.

The background for their approach revolves around some elegant work done on the molecule FK506, a drug used clinically to suppress the immune system. Inside the cell, FK506 targets specific binding proteins—one of which is FKBP12. Investigators have previously shown that a dimeric analog of FK506, its name doubled arithmetically to FK1012, can bind to two copies of FKBP12, producing a dimer. By fusing FKBP12 to another protein, FK1012 can be employed to generate a dimeric form of the hybrid protein. If the protein's new dimer form has a biological property that is different from its monomer form, this property can then be regulated by the concentration of FK1012. In fact, experiments already have shown that with a little protein engineering, FK1012 can regulate transcription, protein trafficking, and receptor activation.

Blau *et al.* applied the FK1012-dimerization technique to determine whether an exogenous drug could be used to control a cell's ability to divide. The cell line they used to test their system was Ba/F3, a line absolutely dependent on interleukin-3 for growth. Introduction of the erythropoietin receptor (EpoR) into these cells allows them to proliferate in a manner dependent on extracellular erythropoietin (Epo). It is known that Epo works by causing the single chain EpoR to dimerize and activate downstream growth-factor signaling.

Blau *et al.* created a DNA clone to express a modified version of the EpoR protein whose Epo binding domain was replaced by three copies of the ligand binding domain of FKBP12 (plus a signal to anchor it intracellularly into the plasma membrane). Introduction of the EpoR–FKBP12 hybrid into the Ba/F3 line yielded clones that could grow only when FK1012 was present in the medium. Cell division could be stopped abruptly by removing FK1012 or adding excess amounts of the monomeric FK506 to act as a competitor of dimerization. Together, these results create a novel on/off switch to control mitogenic signaling.

One use of the mitogenic switch might be to amplify a rare cell type without permanent immortalization. For example, the authors point out that bone marrow stem cells could be targeted with a hybrid receptor and expanded by adding FK1012 ex vivo. After an appropriate period, the drug could be removed and the cells introduced into a host animal or human. The receptor used in these types of studies could be one of the single-chain cytokine receptors such as c-kit.

Another fertile research territory for the FK1012 mitogenic switch might be transgenic mice. One can imagine creating various strains with the hybrid growth factor constructs expressed in different cell types. In this way, one may be able to conditionally clone out undifferentiated cell precursors and allow them to differentiate in vitro with withdrawl of FK1012.

For the future, now that a clever growth switch has been developed, an important step would be to test the positive selection scheme in vivo.

References

1. C. A. Blau *et al.*, *Proc. Natl. Acad. Sci.* **94**, 3076 (1997).

–Robert Sikorski and Richard Peters

Digital Mailbox:
www.sciencemag.org/dmail.cgi?53202a