Instrumentation in the Next Decade

Tomas Hirschfeld

Summary. The progress of instrumentation and measurement science in the next decade will be marked by three major trends. First, as the average instrument achieves a rather considerable level of intelligence, "dumb" systems will become the exception, and we will eventually begin to become proficient in exploiting the resulting capabilities. Second, more sophisticated understanding of measurement science and of actual measurement needs will drive instrumentation design advances such as miniaturized sensors and yet more "hyphenated" instruments and "mapping" instruments. Third, the combination of sensor-based instrumentation and microminiaturization will make possible distributed measurement by allowing point-of-use measurements by nonexperts.

Measurements are the central feature of laboratory science and its applied offshoots in industrial, clinical, environmental, and military practice. And, for better or worse, modern measurement science is dominated by instrumentation. Continued rapid advances in instrumentation and in the methodology of its use are clear megatrends in science and technology.

Intelligent Instruments

In the area of intelligent instruments, the triumphs of microelectronics have been the driving force of a virtual revolution (*1*). With computer power increasing and prices for it decreasing, almost all instruments costing more than $10,000 now contain built-in intelligence, and this threshold is rapidly dropping. Many well-known justifications exist for this (*1*): higher working speed, lower man-hour requirements, added consistency and accuracy, potential for unattended operation, reduction of dexterity factors and transcription errors, built-in logging capability, and, last but most important, flexible and extensive data processing and display. Less well reported, but equally important in the long run, are a number of psychosocial factors such as the computer mystique, the ability of cathode ray tube displays to impress visiting management, the perceived value of obtaining a general purpose computer for other uses, the ability of manufacturers to drastically change their pricing structure when computerizing systems, and keeping up with the Joneses.

As computers become the rule rather than the exception, it becomes more common to equip them with instruments built from the ground up rather than with existing instruments added as attachments. Such computer-based instruments have a number of further advantages of their own. These include cost-saving synergisms in which many instrument components are replaced by software, control panel simplification through software control, the ability to store instrument settings for specific applications as so-called canned expertise, the use of interactive control for data-driven instrument setting and resetting in adaptive operation, and reduction of obsolescence by updating an increasingly dominant software component.

The explosive growth of personal computers will have a strong influence on instrumentation (*2*). On the one hand, their economies of scale provide more hardware for a given cost than microcomputers or larger centralized processors (an idea whose time has come and long gone). On the other hand, the setting of de facto standards for hardware and especially software by the commercial realities of personal computers is now eliminating many of the Tower of Babel problems of laboratory computerization. Former limitations thought to be associated with using personal computers in the laboratory are fast disappearing as the needs of other markets force growth in the power of personal hardware and software. Considering the expected lifetime of current instruments, it is probably already a mistake to design instruments based on anything but personal computers (or personal calculators for low-end applications) in almost all situations.

There is another way in which personal computers will impact the instrumentation field. The physical and accounting lifetimes of most personal computers far exceed their obsolescence lifetimes. We thus face a short-term need for finding retirement homes for millions of used personal computers. The lower computational demands of dedicated simple instruments are an ideal match for these still powerful machines. Add-on hardware to provide such computers with a second career as instruments will be a booming subfield in instrumentation for many years.

To the steady improvement of the computer's capability, we must also add some major technological advances. Here, we must consider small-array processor plug-in boards capable of boosting computational speeds by an order of magnitude; large optical disks, which, at nearly negligible cost, add gigabytes of storage to a personal computer (1 gigabyte of data = 500,000 pages of information); the use of modems, not only for maintenance but also for computer diagnostics (and eventually repair); high-resolution color displays; and voice input/output (I/O).

Clearly, such advances are so large that other parts of the system will promptly become limiting. When a computerized instrument can store the equivalent of a large library in a dozen compact disks, the nonexistence of the library will cause some unhappiness. The efficient generation of large databases in computer-readable media (other than merely bibliographical ones) and new administrative machinery to make these affordable to all will be one of the critical needs of the next few years.

The question of how beneficial such instrument automation may be depends largely on our ability to efficiently use the computer. Except for the very simplest applications, efficient use is a matter of software. To the extent that this requires programming, software depends mainly on our choice of computer language. While assembly programming is largely the province of masochists, the shortcomings of the venerable BASIC and FORTRAN languages are largely compensated for by their high familiarity to users and their simplicity of use. The

Tomas Hirschfeld is a senior scientist at the Lawrence Livermore National Laboratory, Livermore, California 94550

opposite is true for new languages such as FORTH or PASCAL, which have some intrinsic advantages but less widespread use.

A newer trend in software is the use of yet higher level languages, called metalanguages, that are tailored to specific fields of research and use a syntax and command structure corresponding to the normal working procedures in each field. A great many such higher level structures are available for various fields, but their usability is quite uneven. The major problem here is the difference in computer sophistication between programmers and the average end user; the more extensive use of off-site testing by instrumentation manufacturers with naive users will be as essential for scientific software as it has already proven to be for accounting software.

Both the cost and usability of software will be affected by the current expansion of the field. This expansion will make possible both more extensive program preparation efforts and less expensive software. The conventional wisdom that software costs as much as hardware is no longer applicable to software produced for tens of thousands of users; it is evident that lower cost software is around the corner, either through added competition or piracy.

New Data Reduction Procedures

The ongoing expansion in the capabilities of intelligent instrumentation is not yet being matched by a comparable expansion of our ability to exploit these capabilities. The preeminent functions of instrumentation computers are still to measure and to accomplish such elementary data processing steps as signal averaging, smoothing, derivatizing, scaling, linearizing, comparison and substraction, and display controlling. Probably the most sophisticated general applications of computers nowadays involve library searching for qualitative analysis and Fourier transformation as part of the various Fourier transform (FT) spectroscopy techniques. It is noteworthy that each of these new applications has been essential to the development of entire new fields of instrumental work [gas chromatography–mass spectroscopy, FT-infrared (FT-IR), FT–nuclear magnetic resonance].

Over the next decade, many advanced techniques for using instrumentation computers will be put into general use. Probably the most generally valuable of these would be to use the computer to cope with instrumental limitations, errors, and defects (*3*). In the past, huge efforts have been made in removing or minimizing such faults in instrumentation—faults that often involved penalties in cost or performance. Computer correction of these errors is thus a superior alternative for errors that are systematic and known or calculable from the data.

Given the high measurement ratio of signal to noise in modern instruments, as well as the instruments' available computational power, there is no excuse for leaving any such known or calculable errors in the data. A systematic approach to performance improvement in modern instrumentation thus starts with a methodical hunt for problems. Once located, they are studied until their quantitative behavior is understood. In this process, computer modeling studies using synthetic data are quite helpful. Finally, the necessary correction algorithms are developed. Eventually, this will give a lowered residual error consisting of unknown, nonunderstood, or random errors (*3*).

Another major application of computers in instrumentation is the cluster of techniques known as factor analysis, or principal component analysis (*4*). In such analysis, systematic features of a set of experimental patterns (spectra, response curves, or any greater than two-dimensional data fields) are mathematically analyzed to separate out components of these patterns attributable to individual measured sample constituents or individual observed phenomena. This mathematical means of unscrambling measurements owing to various overlapping sources is quite demanding in instrumental signal-to-noise ratio, computer power, and user sophistication (the current bottleneck). Rapid progress in this area can be expected over the next few years.

Another technique for extracting information from the data is the clustering algorithm, in which complex sets of data fields are displayed in such a way that patterns emerge from which generic classification, recognition of relationships, and possible cross-correlations can be detected (*5*). We can also think here of various advances in optimization algorithms, going from the current Simplex algorithm to the recently developed Karmarkar algorithm, which will be in extensive use not only for improving the measurement process but also for improving guidance of the system of which the measurement is a part (*6*).

The common feature of all these advances in computerized instrumentation is the recognition that instrumentation gives us data on measurable parameters, whereas our need is for specific answers to questions about the system being measured.

Modern instruments can shoot out data at a mile a minute at tremendous (and usually hard to visualize) reproducibilities. Unless methods of comparable power are used to convert data to answers, modern instrumentation can easily become a part of the problem rather than of the solution. (Consider the number of volumes in a report submitted to the Food and Drug Administration on the testing of just one new drug: a total of 1200.)

A systematic approach to the transition from data to answers is found in the new discipline of chemometrics (a name that unfortunately implies a nonexistent limitation to chemical problems) (*7, 8*). In this discipline, the relationship between data and information sought is explored as a problem in mathematics and statistics. One of the most useful insights provided by chemometrics is the realization that a cluster of measurements of quantities only distantly related to the actual information sought can be used in combination to determine the information inferentially. Thus, for example, a combination of vapor pressure, refractive index, density, and viscosity can be used to calculate the chemical composition of a solvent mixture. The seeming complexity of such a procedure is often more than compensated for because these individual measurements are either far easier to obtain or often the only ones possible.

As these chemometric procedures grow in complexity, they begin to encroach on what has been called, with infuriating vagueness, artificial intelligence. The dramatic connotation of human-like intelligence is not only currently out of reach, but quite unnecessary. One should reflect here on the complexity of measurement tasks (among others) that can be accomplished by a simple bee. All this is done with a nervous system containing approximately 10^5 neurons, which is considerably less computational hardware than a personal computer. Clearly, then, quite intelligent instrumental operation is possible with the computers of today. Furthermore, it would seem that specialists in measurement science would do well to start talking to entomologists.

A measure of the power of current approaches to intelligent instrumentation is represented by near-infrared reflectance analysis (NIRA) (*9*). This analytical technique attempts to determine the composition of fairly complex samples (such as agricultural products) through near infrared spectroscopy. Here, the

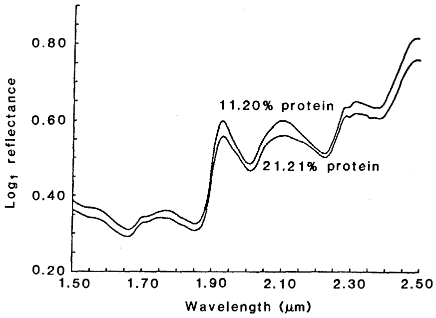


Fig. 1. Spectral differences observed by NIRA.

spectral signature of the various constituents is inaccurately known and heavily overlapped, and the extreme values of composition correspond to barely observable spectral differences, as seen in Figs. 1 and 2.

Instead of conventional data reduction, NIRA uses a combination of spectral correlation and a self-learning algorithm. This operates by measuring a set of preanalyzed samples, cross-correlating the spectra to composition by multilinear regression, and using an optimization algorithm to select a set of measurement wavelengths and calibration coefficients for doing the analysis. (This is then verified with a second, independent, preanalyzed sample set.) This approach provides approximately 0.1 percent repeatable analysis using data in which a 10 percent compositional variation is barely observable to the eye. This power of such self-optimizing procedures using the entire data set goes much further than merely performing difficult measurements. A number of applications of the technique have been successful in determining sample composition for which no known spectral feature, or no spectral feature at all, existed. Here, the algorithm autonomously picked unsuspected spectral phenomena, or cross-correlations with other measurable sample components, to derive a successful measurement method from the learning sample set.

The method is not limited to chemical composition. If, instead, we measure physical sample properties for the learning set, the algorithm comes up with successful methods for measuring these properties. Of course, such advanced data reduction methods are not unique to near infrared spectroscopy. They do require instrumental measurements in which each sample constituent can generate a complex signature, the overall signature can be very accurately measured, and the contributions from each constituent are additive (preferably linearly additive to keep the computer hardware affordable). We will be seeing much more of this approach in future years.

Beyond these ever more sophisticated applications of intelligence to instruments, however, there are some such applications that are different in kind. Once the computer has become a part of the instrument, its capabilities as a controller allow it to take over functions from not only the instrument operator but also the laboratory supervisor. Among these functions are the computer's self-diagnostics (and, given redundance, self-repair); adaptive instrument operation; experiment sequencing and resetting in combination with automatic samplers for unattended operation; annotation where necessary, with coercive support ("I'll display the data if you type in the name of the sample"); compilation of instrument-operating patterns; programmed operator teaching (and testing); and looking for patterns in measurement sets.

Some New Problems

Advances in instrumentation have created a number of new problems. As ever more data are generated in progressively faster measurements, the I/O bandwidth of the human observer is rapidly being exceeded. The signal-to-noise ratio of most measurements exceeds the resolution dynamic range of most displays and the human eye as well (to accurately display data from some modern instruments would require a chart about 60 meters high). The volume of information displayed, even with three-dimensional color motion displays, is slightly ahead of human visual input capabilities and way ahead of human visual memory retention capabilities.

In the long run, ever higher levels of data processing to extract relevant information from the data will be essential. In the medium term, however, more effort in display psychophysics will be needed to match instrument output to brain input. This includes such things as optimized three-dimensional displays by contour mapping and hidden line suppression, nonlinear scaling, difference displays, filled-area false color displays to exploit the area contrast detection capability of the eye, zoom and loupe variable-resolution displays, and oculometer-type displays in which high-resolution information is presented only to the fovea by way of an eye-pointing detector. Instrumentation technology has much to learn in this case from pioneering efforts in aircraft cockpit design.

As computer hardware and software

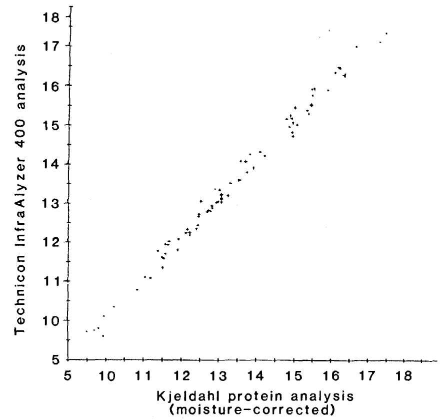


Fig. 2. Quantitative accuracy and reproducibility of NIRA: protein in wheat (77 samples).

become a bigger part of modern instrumentation, they all account for a larger fraction of the flaws associated with instrumentation. Fortunately, computer hardware is already highly reliable, and it is becoming more so as a result of built-in redundancy and self-diagnostics, as well as the recent development of further diagnostics and service available by telephone line.

The situation is otherwise with software, which really brings home the fact that the average computer is really a very fast moron capable of 2 million mistakes per second (*10*). On the one hand, current computer programs are large and complex enough to have a finite probability of error (in addition to the possible error in the basic goal or the algorithm employed). On the other hand, producing new types of data (often not otherwise obtainable) through the use of new and complex procedures usually disables all of our backlog of expertise, problem detection, and error recognition. Modern software, just like its more primitive versions, can on occasion generate nonsense, but now it often is plausible nonsense. When this cannot be otherwise cross-checked, the user is in trouble. In the best of all possible worlds, multiple independent algorithms and programs could be used to verify each other, but this is scarcely practical now.

The often-heard admonition that "to use algorithm XYZ, you must know what you are doing" unfortunately translates as "this method can be believed when it confirms your prejudices," a nonoperative statement. A more appropriate statement would be "a method is useful only if it can be simultaneously surprising and believable." To meet such a requirement, new algorithms must be tested, against independently obtained real data if possible, or against synthetic test data if not.

This widely recognized truth is usually flawed in execution. Experience with mathematical algorithms reveals that, while "bug" frequencies decrease asymptotically with time, they never quite reach zero. In fact, errors are often found in widely used algorithms after years of use because they either require unusual conditions to happen or are hard to detect when they do. Clearly, we need improved software testing methods. The obvious approach here is to go back to the computer that caused the trouble in the first place and write an "exciser" testing program (Fig. 3). This generates entire (and extensive) families of testing data, percolates them through the algorithm to be tested after adding appropriate noise and measurement errors, and checks the end results against their calculated nominal values. Finally, it generates an error report that describes the performance and operating envelope of the algorithm. Over the next decade the operation of such test programs will become a mandatory feature of commercial software development, scientific publication, and new data processing schemes.

Mapping Instruments

The study of organized structures has become a major field of measurement science, driven by the needs of microelectronic technology, biomedical research, and materials science. For this purpose, specialized instrumentation has been developed that combines small sensitive volumes, rapid measurement, spatial scanning, and data storage in image form (*11*).

The most common microprobes illuminate the sample with an electron, x-ray, or ion beam and observe signals produced by such processes as absorption, transmission, scatter, energy loss, photoelectric emission, cathodofluorescence, ion spalling, acoustic surface vibrations, or current pickup. Continued improvements in these microprobes will involve higher spatial resolution (using improved luminosity beam sources), new interaction modes with the sample (including, for example, optical or electrical tunneling, fluorescent energy transfer, microwave emission, and thermal emission), direct storage of images in a computer for more extensive processing, and the combined use of a large number of operating modes to gain more information on the sample (possibly through chemometric interpretation schemes).

A different family of microprobes, which is only now becoming popular, uses optical beams to excite the sample. For the specific case of ultraviolet visible and fluorescence spectroscopies, this usually involves a slightly different technology, in which the whole observation area is illuminated and observed with a stationary television camera. The considerable general utility of such a technique has been held back by the high cost, lack of turnkey quantitative photometry, and insufficient human engineering of available systems, which has kept most users from going beyond mainly qualitative visual microscopy. This situation will change in the near future.

Some new optical microprobes, the FT-IR and Raman microprobes, open up entirely new fields of work because of their ability to provide molecular composition, as well as the atomic composition or physical properties provided by earlier probes. So far, both probes are quite slow and lack much essential software and turnkey features. Nevertheless, these problems can be expected to be solved quite soon.

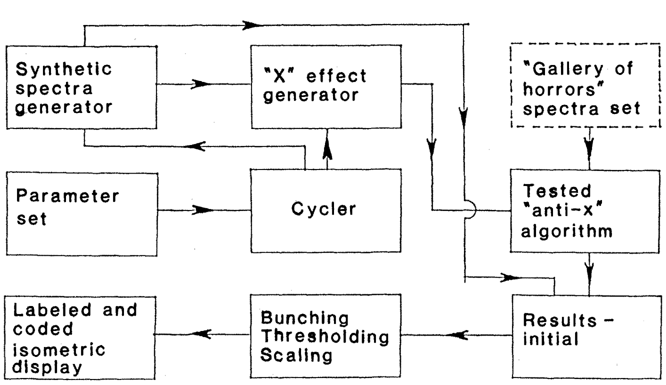
Hyphenated Instruments

Continual increases in sample complexity, measurement difficulty, the amount of information required, and the speed with which the answer is required have created tremendous problems for measurement science. While every one of these requirements can be met by one or another instrument, there is no one instrument that can solve, for example, the problem of doing a rapid, first-run qualitative and quantitative analysis of a complex, totally unknown sample.

However, if one instrument cannot perform this task, a combination of them can. Thus, we come to hyphenated instruments, which can be defined as the marriage of two different instruments of complementary capabilities by way of a sampling interface and a common computer (*12*).

Figure 4 represents a matrix of all the instruments that could be combined to provide a hyphenated instrument. The full squares identify techniques already in use and the half-filled ones identify systems that are potentially feasible but have not yet been built. Many of the advances in instrumentation over the next decade will involve completing this list.

Among the hyphenated instruments



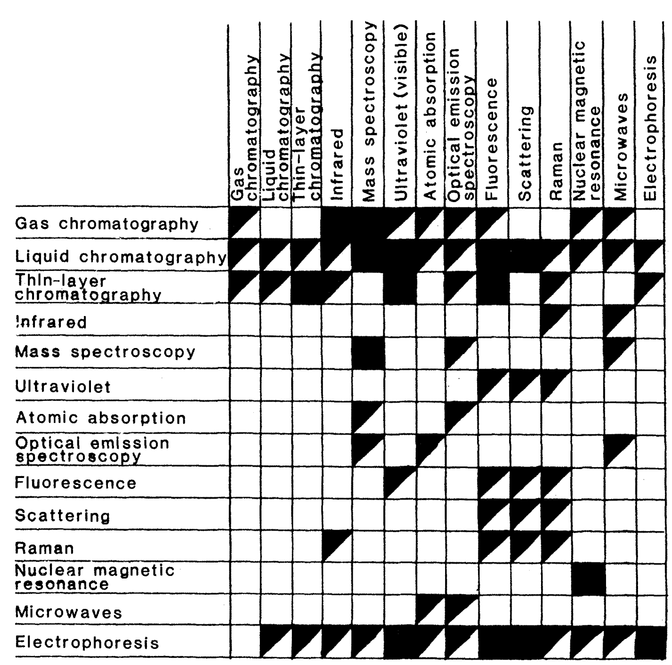


Fig. 3 (left). Block diagram of exerciser software for algorithm testing. Fig. 4 (right). Possible hyphenated methods.

that have already been demonstrated, some have been so successful that they can be expected to be major commercial instruments in the near future. These include gas chromatography–ultraviolet, liquid chromatography–nuclear magnetic resonance, and gas chromatography–gas chromatography. Among other techniques that will become popular in several specialized areas are electrophoresis–mass spectroscopy, particularly suited for biological solutions; gas chromatography–optical emission spectroscopy, a low-cost alternative to gas chromatography–mass spectroscopy (GC-MS); electrophoresis-fluorescence, a promising combination of flow photometric with capillary electrophoresis technology, again optimal for biological solutions; and electrophoresis–liquid chromatography, an outgrowth of the convergent evolution of liquid chromatography and electrophoresis as column diameters become smaller.

Of course, the list of possibilities is growing rapidly, and so we must consider both supercritical fluid chromatography and flow injection assay as further candidate components for hyphenated instruments in the near future. Here, we must also consider ternary systems, of which GC-MS-IR is currently the one closest to general availability and will probably be the first one suitable for dealing with the problem of the completely unknown sample.

The amount of data that hyphenated systems can generate is immense, and this fact imposes serious hardware and software constraints on the system. A more serious concern is that these systems also require enormous databases for rapid automatic data interpretation. And, most serious of all, these enormous databases are not even close to being available. Despite all that can and should be done in this area, we should not expect these problems to be solved in the next decade or even beyond it.

Fortunately, an alternative to these enormous databases in automatic data interpretation is possible. Basically, the need for them arises because most of our qualitative analysis procedures operate by recognition against databases, which therefore need to contain information about hundreds of thousands of compounds. In another approach software is used to make molecular structure determinations from the data through fundamental physical chemistry and spectroscopy. This allows qualitative interpretation to become mostly independent of large databases. The difficulty here lies in the extreme complexity of the programming, which definitely belongs in the advanced artificial intelligence field. Very active efforts in this field at a number of locations, however, make the prospects good for an advance in this area in the next 5 years (*13, 14*).

Sensors, Miniaturization, and Distributed Instrumentation

As more and more of the functions of modern instrumentation are taken over by electronics, the field as a whole is rapidly evolving toward sensor-based instrumentation design (*15*). This involves a generalized electronics system, more often than not a microcomputer with appropriate I/O devices and software, that is coupled to a system of sensors that transduce the external measurable quantities into electrical signals.

The off-loading of the largest possible portion of the sensor tasks onto the data system is then used to simplify the former. Consequently, it is possible, for example, to use self-calibration loops in place of reproducibility or stability; nonlinear processing instead of linear sensor response; adaptive system settings in place of dynamic range; correction using internal or external information in place of freedom from perturbations; multiple sensors acting cooperatively with pattern recognition in lieu of selectivity; time-dependent and circumstance-dependent correction algorithms rather than freedom from aging; and digital time filtering instead of a well-behaved time response. In effect, we are replacing nearly all the charcteristics formerly thought necessary in an analytical system by enough signal-to-noise ratio to do corrections, with a system uniform and well understood enough so that we know how to do them, and with the processing power with which to do them.

In the past, there have been impressive examples of the power of this approach in, for example, FT-IR. In this instance, system construction tolerances of the order of an atomic diameter were replaced by computer feedback loops to generate simple, low-cost, and even field-portable instruments.

The cost-effectiveness of such sensor-based instrumentation is quite impressive. Here, the major remaining part of the instrument, the data system, benefits from all the cost benefits of commonality, large production volumes, and microelectronic revolution. This evolution of a major part of the instrumentation field into sensor peripherals of personal computers is not new; nature has been doing nothing else for a few hundred million years.

The design and development of sensors will be one of the major instrumentation research fields of the next decade, and scarcely any physical or chemical phenomenon will be left unused for this purpose. This includes many phenomena whose applicability to the measurement at hand may at first appear dubious, in that the power of chemometrics (and parallel disciplines in other areas) will be used to arrive indirectly at the desired data.

In parallel with this research, much effort will be devoted to the miniaturization of sensors. For this, again, it will be possible to use microelectronic fabrication technology (a set of tools far too good to leave for the electronic engineers alone). Here, however, the need for invention quickly arises.

The laws of scaling, of which the square-cube law of structural engineering is the best-known example, stipulate that small devices cannot be built by merely scaling down small ones. The interrelationship of physicochemical phenomena with the distance scale becomes quite dramatic as device sizes shrink toward the micron scale.

Thus, we see viscous effects dominating inertial ones in fluid dynamics, and being superseded by Brownian forces in their turn; electrostatics becoming preeminent over electromagnetics; quantum effects becoming major features of the system; surface forces acquiring major significance; and so forth. Obviously, instrument and sensor design must become very different under these circumstances. Fortunately, a major shortcut is possible for research in this area: plagiarism from the world of nature, which has traditionally designed its sensors in precisely this size scale (*16*).

The value of such miniaturization arises from the simultaneous miniaturization of the electronics, in that it allows in succession, portable, then hip-pocket, and then point-of-use implanted instrumentation. This reduction in size, already attained by some of the simpler electronic measurement instruments, will become a general feature of all kinds of instrumentation in the near future.

The significance of this evolution can be appreciated by drawing a parallel with the computer world. Instrumentation, particularly in the chemical and biological field, is in much the same condition as computing was 15 years ago. We have large mainframes at a distant location, operated by a mysterious priesthood that kindly (but obscurely) explains to us today what was the matter with yesterday's samples, and then lets us try again. The sense of power people feel when acquiring pocket calculators and then personal computers can be appreciated only by those old enough to still remember owning mechanical slide rules.

Precisely the same transition is at hand in chemical and biological laboratories, where miniaturization and machine intelligence will return measurement power to the end user, while freeing him from becoming either a specialist in a different field or an almost-full-time money-raiser to do research in his own field. This open-ended task will occupy the coming decade and more, but will be indeed worth the effort.

References and Notes

1. S. Heller and R. Poterzone, Eds., *Computer Applications in Chemistry* (Elsevier, New York, 1983).
2. P. Lykos, *Personal Computers in Chemistry* (Wiley-Interscience, New York, 1981).
3. T. Hirschfeld, in *Fourier Transform Infrared Spectroscopy*, J. Ferraro, Ed. (Academic Press, New York, 1979), p. 193.
4. E. R. Malinovski and D. G. Howery, *Factor Analysis in Chemistry* (Wiley-Interscience, New York, 1980).
5. D. L. Massart and L. Kaufman, *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis* (Wiley, New York, 1983).
6. W. E. Biles, *Optimization and Industry Experimentation* (Wiley-Interscience, New York, 1980).
7. B. Kowalski, *Chemometrics: Theory and Application* (American Chemical Society, Washington, D.C., 1977).
8. M. Delaney, *Anal. Chem.* **56**, 261 R (1984).
9. T. Hirschfeld and E. Stark, in *Food Analysis*, G. Charambolous, Ed. (Wiley-Interscience, New York, 1984), p. 505.
10. T. Hirschfeld, paper presented at the American Society for Testing Materials Symposium on Computerized Infrared, Philadelphia (1984).
11. G. A. Kentgen, *Anal. Chem.* **56**, 69 R (1984).
12. T. Hirschfeld, *ibid.* **52**, 297 A (1980).
13. P. C. Jurs and T. L. Isenhour, *Chemical Applications of Pattern Recognition* (Wiley-Interscience, New York, 1975).
14. K. Eckschlager and V. Stepanek, *Information Theory as Applied to Chemical Analysis* (Wiley-Interscience, New York, 1984).
15. T. Hirschfeld, J. Callis, B. Kowalski, *Science* **226**, 312 (1984).
16. T. Hirschfeld, *Energy Technolgoy Review* (Lawrence Livermore National Laboratory, Livermore, Calif., February 1984), pp. 16–23.
17. This work was performed under the auspices of the Department of Energy by the Lawrence Livermore National Laboratory under contract W-7405-ENG-48.

RESEARCH ARTICLE

Molecular Cloning of a Complementary DNA Encoding Human Macrophage-Specific Colony-Stimulating Factor (CSF-1)

Ernest S. Kawasaki, Martha B. Ladner, Alice M. Wang
Janelle Van Arsdell, M. Kim Warren, Mazie Y. Coyne
Vicki L. Schweickart, Mei-Ting Lee, Kenneth J. Wilson
Albert Boosman, E. Richard Stanley, Peter Ralph, David F. Mark

The generation of granulocytes and macrophages from immature hematopoietic progenitor cells depends on the presence of several hormone-like, glycoprotein growth factors, the colony-stimulating factors or CSF's (*1*). There are several subclasses of CSF's, and they can be distinguished by the kind of mature cells that they produce in semisolid culture media.

One subclass, CSF-1, stimulates monocyte and macrophage production predominantly or exclusively (*2*). Other human CSF's stimulate the proliferation of progenitors committed to neutrophil and macrophage, and neutrophil or eosinophil lineages (*3*). Interleukin-3 (IL-3) (*4, 5*) in the murine system and, perhaps, a similar pluripotent CSF in the human (*6*) can stimulate the proliferation of monocytic, granulocytic, erythroid, megakaryocytic, and mast cell colonies. Although all these CSF's, with the exception of the eosinophilic type, have been purified, the small quantities available from natural sources have hindered their biochemical and biological characterization. The molecular cloning of two of the above factors, murine IL-3 (*7, 8*), and murine and human granulocyte-macrophage CSF (GM-CSF) (*9, 10*), should aid in elucidating the mechanisms by which these hematopoietic factors act.

Native human CSF-1 is a heavily glycosylated homodimer with a molecular size of ~45,000 daltons (*11*). It can be clearly distinguished from the other subclasses by specific radioimmuno- and radioreceptor assays (*2*). The proliferative effects of CSF-1 are restricted solely to the mononuclear phagocytic lineage (*12*), and the specific cell receptors that mediate the biological effects of CSF-1 apparently occur exclusively on cells of the same lineage (*13*). In the murine system, L cell–conditioned medium is a convenient source of CSF-1 (*14*), and amino-terminal and internal amino acid sequences have been determined (*15*).

Isolation of CSF-1 genomic clones. We now describe the isolation of CSF-1 genomic clones using oligonucleotide probes derived from the amino terminal sequence of human CSF-1. Human urine was used as a source of CSF-1 protein (*2*). CSF-1 was purified as described (see legend to Fig. 1), and the NH_2-terminal sequence was determined (*16*). The 12–amino acid NH_2-terminal sequence obtained from the human urinary CSF-1 (Fig. 1A) is highly homologous to the NH_2-terminus of murine L-cell CSF-1 (see Fig. 2B). The finding that the NH_2-termini were similar in sequence indicated that we were studying the same protein, even though the CSF-1 was prepared from widely different sources. Two oligonucleotide probes were derived from the NH_2-terminal sequence (Fig. 1B). A 16-fold degenerate, 35-base oligonucleotide, complementary to the NH_2-terminus, was constructed according to preferred mammalian codon usage (*17*). An 18-base, 64-fold degenerate oligonucleotide derived from amino acids 5 through 10 was also made to aid in screening.

Using the two oligonucleotides described above, we screened a human genomic library (*18*), and several CSF-1 genomic clones were isolated. Restriction enzyme and Southern gel (*19*) analysis of nine separate phage isolates showed that a common 3.8-kilobase (kb) Hind III fragment hybridized strongly to both of the oligonucleotide probes. This Hind III fragment was subcloned into M13mp19 (*20*) for further analysis (Fig. 2A). DNA sequence analysis (Fig. 2B)

E. S. Kawasaki, M. B. Ladner, A. M. Wang, M. Y. Coyne, V. L. Schweickart, and D. F. Mark are in the Department of Molecular Biology; J. Van Arsdell is in the Department of Microbial Genetics; M. K. Warren, M-T. Lee, and P. Ralph are in the Department of Cell Biology; and K. J. Wilson and A. Boosman are in the Department of Protein Chemistry at Cetus Corporation, 1400 Fifty-Third Street, Emeryville, California 94608. E. R. Stanley is in the Departments of Microbiology and Immunology, and of Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461. Send correspondence to E.S.K. The present address for K.J.W. is Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404.