This PDF is available from The National Academies Press at http://www.nap.edu/catalog.php?record_id=19410

FROM THE ARCHIVES

Research Briefings, 1984 (1984)

Pages
128

Size
8.5 x 10

ISBN
0309034957

Committee on Science, Engineering, and Public Policy; National Academy of Sciences; National Academy of Engineering; Institute of Medicine

Find Similar Titles

More Information

Visit the National Academies Press online and register for...

- Instant access to free PDF downloads of titles from the
  - NATIONAL ACADEMY OF SCIENCES
  - NATIONAL ACADEMY OF ENGINEERING
  - INSTITUTE OF MEDICINE
  - NATIONAL RESEARCH COUNCIL
- 10% off print titles
- Custom notification of new releases in your field of interest
- Special offers and discounts

Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

To request permission to reprint or otherwise distribute portions of this publication contact our Customer Service Department at 800-624-6242.

Copyright © National Academy of Sciences. All rights reserved.

NATIONAL ACADEMY OF SCIENCES
1863–2013
Celebrating 150 Years of Service to the Nation

RESEARCH BRIEFINGS 1984

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

RESEARCH BRIEFINGS 1984

for the Office of Science and Technology Policy,
the National Science Foundation,
and Selected Federal Departments and Agencies

Committee on Science, Engineering,
and Public Policy
National Academy of Sciences
National Academy of Engineering
Institute of Medicine

NAS-NAE
NOV 1 1984
LIBRARY

NATIONAL ACADEMY PRESS
Washington, D.C. 1984

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Q
127
.U6
C63
1984
c.1

National Academy Press 2101 Constitution Avenue, NW Washington, DC 20418

NOTICE: The National Academy of Sciences was established in 1863 by Act of Congress as a private, nonprofit, self-governing membership corporation for the furtherance of science and technology for the general welfare. The terms of its charter require the National Academy of Sciences to advise the federal government upon request within its fields of competence. Under this corporate charter, the National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively.

The Committee on Science, Engineering, and Public Policy is a joint committee of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. It includes members of the councils of all three bodies.

This work was supported by the National Science Foundation under Contract PSPF91304.

Library of Congress Catalog Card Number 83-63234

International Standard Book Number 0-309-03495-7

Copyright © 1984 by the National Academy of Sciences

No part of this book may be reproduced by any mechanical, photographic, or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted, or otherwise copied for public or private use, without written permission from the publisher, except for the purposes of official use by the United States Government.

Printed in the United States of America

Copyright © National Academy of Sciences. All rights reserved.

OCLC# 12215872

JUN 24 1986

Committee on Science, Engineering, and Public Policy

LEON T. SILVER, Professor of Geology, Division of Geological and Planetary Sciences, California Institute of Technology (*Chairman*)
LINDA H. AIKEN, Vice President for Research, The Robert Wood Johnson Foundation
RICHARD C. ATKINSON, Chancellor, University of California, San Diego
JACOB BIGELEISEN, Leading Professor, Department of Chemistry, State University of New York, Stony Brook
FLOYD E. BLOOM, Director and Member, Division of Pre-Clinical Neuroscience and Endocrinology, Scripps Clinic and Research Foundation
W. DALE COMPTON, Vice President-Research, Ford Motor Company
EMILIO Q. DADDARIO, Wilkes, Artis, Hedrick and Lane, Attorneys at Law
GERALD P. DINNEEN, Vice President, Science and Technology, Honeywell, Inc.
ALEXANDER FLAX, President Emeritus, Institute for Defense Analyses
GARDNER LINDZEY, President and Director, Center for Advanced Study in the Behavioral Sciences
†GEORGE M. LOW, President, Rensselaer Polytechnic Institute
EDWARD A. MASON, Vice President-Research, Standard Oil Company (Indiana)
*JOHN L. McLUCAS, Executive Vice President and Chief Strategic Officer, Communications Satellite Corporation
*ELIZABETH C. MILLER, Van Rensselaer Potter Professor of Oncology, McArdle Laboratory for Cancer Research, University of Wisconsin
DANIEL NATHANS, Professor, Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine
GILBERT S. OMENN, Dean, School of Public Health and Community Medicine, University of Washington, Seattle
HERBERT A. SIMON, Professor of Computer Science and Psychology, Department of Psychology, Carnegie-Mellon University

†Past Chairman, deceased July 17, 1984
*Term expired June 30, 1984

Copyright © National Academy of Sciences. All rights reserved.

*I. M. SINGER, Professor, Mathematics Department, University of California, Berkeley
F. KARL WILLENBROCK, Cecil H. Green Professor of Engineering, School of Engineering and Applied Science, Southern Methodist University

Ex Officio

FRANK PRESS, President, National Academy of Sciences
ROBERT M. WHITE, President, National Academy of Engineering
FREDERICK C. ROBBINS, President, Institute of Medicine

COSEPUP Staff

ALLAN R. HOFFMAN, *Executive Director*
BARBARA A. CANDLAND, *Administrative Assistant*

*Term expired June 30, 1984

Copyright © National Academy of Sciences. All rights reserved.

Preface

This volume contains nine research briefing reports prepared in 1984 under the supervision of the Committee on Science, Engineering, and Public Policy (COSEPUP), a joint committee of the National Academies of Sciences and Engineering, and the Institute of Medicine. Prepared at the request of the Office of Science and Technology Policy (OSTP) and the National Science Foundation (NSF), the briefings were delivered in both written and oral form to these organizations between September 5 and October 1, 1984. The briefings were also presented to senior officials of other interested federal departments and agencies.

Potential briefing topics were identified by OSTP, the Academies, the Institute of Medicine, and units of the National Research Council. Final selections were made by OSTP. It must be emphasized that presentation of these nine briefings in 1984 by COSEPUP does not imply that other topics not presented are considered to be of lesser importance to the progress of U.S. science and technology. No priority setting among fields of science and technology by COSEPUP is implied. The list reflects the fact that only a limited number of timely topics can be covered in a given round.

The nine briefing topics for 1984 are:

1. *Computer Architecture*: Research directed toward increasing the speed and capabilities of computers through innovative hardware design and software optimization, including parallel processing.

2. *Information Technology in Precollege Education*: Research directed toward identification of special opportunities for use of new information technology in precollege education offered by recent work, notably in the cognitive sciences, artificial intelligence, hardware, and analyses of the strengths and limits of currently available and anticipated devices.

3. *Chemical and Process Engineering for Biotechnology*: Long-range fundamental research directed toward developing new reactor concepts (immobilized-enzyme reactors and continuous large-scale cell culture systems) and developing new unit operations for manipulating and purifying fragile macromolecules in highly complex product mixtures.

4. *High-Peformance Polymer Composites*: Research directed toward synthesis of high-strength, high-modulus polymers and polymer composites for structural applications, including examination of relationships between microstructure and performance within single-phase materials, and relationships between surface characteristics and interfacial properties in multiphase composites.

5. *Biology of Oncogenes*: Research directed toward enhanced understanding of the genetic characteristics of cellular and viral oncogenes, the process of activation of oncogenes and its relationship to chemical/environmental carcinogenesis, and the nature of oncogene products and how they subvert normal orderly cellular growth and development.

6. *Interactions Between Blood and Blood Vessels (Including the Biology of Atherosclerosis)*: Research directed toward enhanced under-

Copyright © National Academy of Sciences. All rights reserved.

standing of the interplay between blood constituents and the blood vessel wall, ranging from molecular and cellular interactions to the role of injury and subsequent disease.

7. *Biology of Parasitism*: Research directed toward enhanced understanding of how parasites establish and maintain relationships with their hosts and the consequent disease processes, including examination of:

- cell membrane phenomena;
- intermediary metabolism;
- life cycle and development changes in vectors and in hosts;
- immune system and other protective responses and their underlying genetic and regulatory mechanisms; and
- related opportunities to develop vaccines, pharmaceuticals, and biological controls.

8. *Solar-Terrestrial Plasma Physics*: Research directed toward understanding the interconnected physical processes responsible for the injection, transport, energization, and loss of plasma from regions on the sun, in the solar winds, and in the earth's magnetosphere.

9. *Selected Opportunities in Physics*: Research directed toward enhanced understanding of the following promising areas of physics: physics at the laser-atomic frontier; relativistic plasma waves; physical properties of deliberately-structed materials; biomolecular dynamics and intercelluar cooperativity; cosmology; and nuclear matter under extreme conditions.

This set brings to 21 the number of research briefings developed for OSTP and NSF by COSEPUP in three separate rounds. Those topics for which briefings were developed in 1982 and 1983 are listed below:

1982 (Published as *Research Briefings for the Office of Science and Technology Policy, the National Science Foundation, and Selected Federal Departments and Agencies*, National Academy Press, 1983).

1. Mathematics
2. Atmospheric Sciences
3. Astronomy and Astrophysics
4. Agricultural Research
5. Neuroscience
6. Human Health Effects of Hazardous Chemical Exposures
7. Materials Science

1983 (Published as *Research Briefings 1983 for the Office of Science and Technology Policy, the National Science Foundation, and Selected Federal Departments and Agencies*, National Academy Press, 1983).

1. Selected Opportunities in Chemistry
2. Cognitive Science and Artificial Intelligence
3. Immunology
4. Solid Earth Sciences
5. Computers in Design and Manufacturing

Each briefing is developed by a panel of approximately 12 members, charged with critically assessing its field, and identifying those research areas within the field likely to return the highest scientific dividends as a result of incremental federal investment in the upcoming fiscal year. Panels generally meet once, for several days, to carry out their charge, with rapporteurs present to summarize the discussions and prepare initial drafts of briefing papers. After further review by the panels, the briefing papers and oral presentations are reviewed by COSEPUP, prior to their delivery to federal officials in three separate briefings. Funds to support this round, as well as both earlier rounds of research briefings, have been provided by the National Science Foundation.

Publication of this volume comes shortly after the untimely death of COSEPUP's past Chairman, George M. Low, under whose capable leadership the research briefing activity was initiated and developed. George's guidance and wisdom will be sorely missed.

Leon T. Silver, *Chairman*
Committee on Science, Engineering, and Public Policy

Copyright © National Academy of Sciences. All rights reserved.

Contents

Report of the Research Briefing Panel on Computer Architecture 1

Report of the Research Briefing Panel on Information Technology in Precollege Education 17

Report of the Research Briefing Panel on Chemical and Process Engineering for Biotechnology 31

Report of the Research Briefing Panel on High-Performance Polymer Composites 45

Report of the Research Briefing Panel on the Biology of Oncogenes 57

Report of the Research Briefing Panel on Interactions Between Blood and Blood Vessels (Including the Biology of Atherosclerosis) 69

Report of the Research Briefing Panel on the Biology of Parasitism 81

Report of the Research Briefing Panel on Solar-Terrestrial Plasma Physics 91

Report of the Research Briefing Panel on Selected Opportunities in Physics 105

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Report of the Research Briefing Panel on Computer Architecture

Copyright © National Academy of Sciences. All rights reserved.

Research Briefing Panel on Computer Architecture

Jacob T. Schwartz (*Chairman*), New York University, New York, N.Y.
Richards Adrion, National Science Foundation, Washington, D.C.
Donald Bonstrom, ETA Systems, St. Paul, Minn.
Charles Bostick, National Security Agency, Fort Meade, Md.
John Cavalini, Department of Energy, Washington, D.C.
Robert J. Douglass, Los Alamos National Laboratory, Los Alamos, N. Mex.
W. Daniel Hillis, Thinking Machines Corporation, Waltham, Mass.
Thomas A. Keenan, National Science Foundation, Washington, D.C.
David Kuck, Kuck and Associates, Champaign, Ill.
H. T. Kung, Carnegie-Mellon University, Pittsburgh, Pa.
Thomas Nash, Fermi National Accelerator Laboratory, Batavia, Ill.
James Pomerene, IBM Corporation, Yorktown Heights, N.Y.
John Riganati, National Bureau of Standards, Gaithersburg, Md.
Arnold Reisman, Microelectronics Center of North Carolina, Research Triangle Park, N.C.
Paul Schneck, Office of Naval Research, Arlington, Va.
Stephen Squires, DARPA/IPTO, Arlington, Va.
Kenneth Wallgren, NASA, Washington, D.C.

Staff

Mark Weiser, *Rapporteur*, University of Maryland, College Park, Md.
Gary E. Clark, *Staff Director*, Computer Science and Technology Board
Rosena A. Ricks, *Administrative Secretary*, Computer Science and Technology Board
Allan R. Hoffman, *Executive Director*, Committee on Science, Engineering, and Public Policy

Copyright © National Academy of Sciences. All rights reserved.

Report of the Research Briefing Panel on Computer Architecture

I. INTRODUCTION

The market for supercomputers is only a small part of the total computer market. However, it represents a critical cutting edge for the advance of computer technology and provides machines that are critical to technological leadership in a whole range of areas varying from aircraft design to basic chemistry. For this reason it is vital that the programs that achieved U.S. supercomputer leadership be continued and that new programs to maintain that leadership be initiated.

Rapidly developing VLSI technology has created the basis for major advances in superspeed computer performance. Maximum attainable computation rates, which are approaching 1,000 MIPS (millions of instructions per second) for the largest vector supercomputers, will rise over the next decade through 20,000 MIPS, and possibly to levels of 100,000 MIPS and more. Attaining such extreme speeds will depend on the effective use of parallelism, i.e., on the development of computers that can execute many hundreds, thousands, or tens of thousands of instructions simultaneously and of software that can effectively orchestrate many simultaneous streams of computation. The ability to explore this essential new opportunity has been opened up by the amazing reductions in circuit costs attained during the past decade.

This document will review the current supercomputer situation and will predict developments over the next 5 to 10 years. The following major areas will be considered:

1. the *chip and mass storage technologies* on which U.S. primacy in the computer area, and its ability to advance, ultimately rest;
2. the *computer architectures*, both novel and relatively conventional, likely to be important for large-scale computing over the next decade;
3. the potentially significant role of *special-purpose computers*, which may be of particular importance for certain types of research (e.g., computer vision, speech and signal processing, scientific data reconstruction and reduction), where dedicated equipment and very high performance, but relatively inexpensive equipment, are all essential;
4. the extent to which *software technology* can and must support the move to large-scale parallel computing;
5. the developments in *communication tech-*

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

nology that will define the ways in which supercomputers can be made available to the U.S. research community;

6. the *efforts abroad* that might undercut the dominant U.S. role in high-performance computing;
7. the relationship of supercomputing to the broader field of *artificial intelligence*; and
8. the bottom-line technical policy question of how to get the U.S. research and industrial communities to *move forward in rapid and effective cooperation*.

II. CHIP TECHNOLOGY

The astonishing exponential growth of microchip performance is both the main driving force for the computer industry and the technical basis for expected parallel supercomputer developments. Though many circuit families of differing speeds and cooling requirements play a role in this growth, the following remarks will concentrate on just one family, the so-called field-effect transistors (FETs), which generate 40 times less heat than the faster bipolar chips, a very significant advantage as the number of circuits per chip increases steadily. State-of-the-art FET chips as of 1984 are represented by the 256K memory chip (512,000 devices), just now coming into mass production, and the roughly 100,000-device Motorola 68000 microprocessor. The 5 to 1 circuit density advantage of memory chips results from their very regular internal geometric structure and highly optimized design. Since 1960 these circuit densities have doubled roughly every 18 months, and this growth is expected to continue, with memory and microprocessor densities remaining in a similar proportion. By 1988 this should make 2-million-bit/chip memories available and could make million-transistor single-chip processors possible. As circuit sizes shrink, speeds will rise, so by 1988 the cycle-time of single-chip FET processors should have fallen from the 80 nanosec now typical to 30 nanosec for chips operated at room temperature and 15 nanosec for chips operated cryogenically. A typical application of this technology might be a single-chip, 64-bit scientific processor with full floating-point capability, on-chip memory management, internal overlap of instruction execution, 32 Kbytes of on-chip cache memory, and the capability of performing 20-40 MIPS. Processors tailored for somewhat faster execution of higher-level languages such as LISP and special-purpose computing devices like those described in Section V below (e.g., specialized vision-oriented or switching chips) all can be supported by this same technology.

The use of new materials such as gallium arsenide and other techniques for attaining higher electron mobility should allow perhaps 10 times higher performance, but they require significantly more difficult process and materials technology. The Japanese HEMT and the U.S. VHSIC are examples of such projects. Josephson circuits, which operate at cryogenic temperatures, and 3-D chips, which try to build up circuits in the third dimension rather than using the essentially planar structures now common, are not likely to be useful technologies in the near future.

Wafer-scale integration, including full and partial wafer techniques, have been considered for reducing the costs associated with the currently used wafer-dicing and reassembly technology. This approach faces severe chip yield problems, and one must reckon with the unfavorable signal-propagation characteristics of long on-chip conductors. For this reason, a packaging technique like that used in the IBM Thermal Conduction Modules, i.e., bonding individual chips to a multilayer ceramic substrate, may dominate for high-performance devices. Assuming that 200-400 chips are bonded to a single 6″ × 6″ ceramic board of this type, two such boards might by 1990 produce a 64-processor, 1,200 MIPS peak-rate parallel computer at a manufacturing cost between $50 thousand and $100 thousand.

Attaining the chip performances described in the preceding paragraphs will require attention to a challenging mass of supporting technologies, including special materials development, x-ray, E-beam, and ion lithography, as well as radiation damage removal

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

methods. This work is moving rapidly in the United States and at least as rapidly in Japan, where experimental and fabrication facilities sometimes bettering the best U.S. facilities have been put in place. Japanese work in this critical area is spurred by a particularly clear national determination to succeed, resulting in a willingness to try approaches that U.S. workers tend to avoid because of initially unclear market perspectives. In this determination, combined with the dedicated large teams and engineering excellence brought to bear, lies the technological roots of the Japanese computer industry's rapid advance.

III. DISK TECHNOLOGIES

For supercomputers to be used effectively, they must be supplied as balanced systems, with input-output (I/O) bandwidth, external storage, displays, and other capabilities all appropriately matched to the expected two- or three-orders of magnitude increase in computation rates. Concern has been expressed about the ability of storage disks, a key element of this "peripheral" technology, to keep up with such rapid speed increases. The workhorse disk for current supercomputer systems is still the CDC Model 819 disk, a rotating storage device dating from 1970, with a capacity of 600 million bytes and a 5-Mbyte-per-second transfer rate. Two newer disks just now becoming available provide roughly double this capacity and storage rate, an exceedingly modest increase compared to far more rapid computation-rate speedups. Other more ambitious magnetic disk alternatives include the possible use of synchronized spindles or multiple heads in parallel. Although no such higher performance products have been announced (probably because of less-than-satisfactory past market experience with very-high-performance disks), products of this type may appear as derivatives of disk and head technology developed for larger, more lucrative markets.

Against this background, the development of a new optical disk storing 10^{11} bytes, now in progress under NASA and USAF (RADC) sponsorship, appears extremely significant. This disk, a first version of which should be available in 1987, promises to support a 10^8 byte/sec transfer, much better matched to anticipated supercomputer requirements than anything currently available. Access time, i.e., the time needed to start the inflow of a randomly selected data record, is anticipated to be 50-100 milliseconds. The development of this potentially important device ought to be pushed rapidly to prevent an initially limited market from delaying its appearance.

IV. CURRENT AND NEW COMPUTER ARCHITECTURES

During the next 5 years, new types of machines based on large-scale parallelism, having the ability to perform thousands or even tens of thousands of arithmetic operations simultaneously, will begin to appear alongside high-speed sequential and vector computers that dominate today's high-end computer market. At a rate determined by their performance, their cost, and their suitability for various applications, these new machines may progressively supplant present designs in some market sectors. Though our panel did not make any precise predictions concerning the rate at which this development will unfold over the next decade, we can describe the main architectural considerations relevant to this process.

High-speed single-thread machines keep programming relatively straightforward by behaving externally as if they execute instructions serially, even though internally a small number of operations (5-20) are actually being handled simultaneously. Machines of this design are the staple of present medium-high performance (4-20 MIPS) computing. It is technically possible to raise the performance of these machines to the 80-150 MIPS range with current technologies and to 250 MIPS or somewhat more using more advanced technology such as that represented by GaAs. Performance improvements beyond this will

Copyright © National Academy of Sciences. All rights reserved.

depend on the parallel use of multiple high-speed serial processors. Very-high-speed serial machines and multiprocessor configurations involving small numbers of such computers are bound to retain major, perhaps dominant, economic importance for the next 5-10 years, especially for commercial applications. In the supercomputer environment, fast (4-10 MIPS) serial computers will be the single-user "workstations" that supply researchers with their software development tools and much of their graphic and data display environments.

Vector supercomputers, like the present U.S. Cyber 205, CRAY XMP, and the Japanese Hitachi S-810/20 and Fujitsu VP-200, can be regarded as simple parallel machines specialized for the handling of computations characterized by particularly regular patterns of data motion. Such regularities typify certain important numerical scientific codes, and hence fast vector machines are generally designed as "scientific number crunchers" equipped with the highest-speed floating-point operation units. The current peak performance of these machines is in the hundreds of MIPS. Single-processor machines of this type could achieve operation rates of 250-500 MIPS with the best present technology and perhaps twice this using GaAs technology. Multiprocessor configurations of vector processors, a type of design under active consideration by several supercomputer manufacturers, will probably attain peak performances of roughly 10,000 MIPS* using 16-32 processors within 5 years, rising to 25,000 MIPS within the coming decade. At such performance levels, they can be expected to retain a major role, and perhaps a dominant role, in bread-and-butter scientific number crunching through much of the decade. Note, however, that computers of this design can exploit parallelism only in innermost program loops; and for programs that branch or address memory in irregular patterns, the actual efficiency typically decreases to 15-20 percent of rated peak performance, e.g., for such problems a 4,000 MIPS actual computation rate is all that can be expected from a machine nominally rated at 25,000 MIPS.

New parallel supercomputer designs, many of which are still highly experimental, are at once the focus of much current research and the one area (of design, as distinct from technology) in which sudden surprising advances might occur. Many of these designs promise large increases in computational power without significant redesign of their components, i.e. they are "scalable." With such architectures it may be possible to increase machine speeds by factors of 10 or even 100 by increasing the amount of parallelism used, without major redesign but at a proportional increase in cost. Once suitable software has been developed for machines of this type, a major research undertaking, the homogeneity of their structure should make it possible to move applications to much larger-scale parallel machines with minimal software change. Moreover, scalable parallel machines have major potential advantages in a VLSI environment because they are built from large numbers of identical parts that can be mass produced efficiently and because the design cost can be amortized over multiple configurations differing only in size. For all these reasons, we review this class of computers at length and attempt to clarify the technical options available for parallel supercomputer development by listing various principal design alternatives characterizing the design of such machines.

Most of the more than 70 parallel machines that have been proposed by university and

*The performance of high-speed scientific computers is often rated not in terms of the total rate at which instructions of all types are performed but only counting that part of the total activity devoted to the execution of "floating-point" arithmetic operations: MFLOPS (millions of floating-point instructions per second) rather than MIPS (millions of instructions per second). The ratio between these two measures is typically taken to be 1/3, but since this ratio is highly application dependent, we have chosen to sidestep this distinction and to state all crude performance figures in MIPS.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

industrial groups here and abroad occupy easily characterized positions in the "design space" defined by the following alternatives:

1. *SIMD versus MIMD.* An SIMD (single-instruction multiple-data) machine consists of multiple processors (or of a few very fast pipelined processing elements), all of which are fed synchronously by a single instruction stream but which operate on independent streams of data. Vector machines like the CRAY I or Hitachi 810, and synchronous array machines like the ILLIAC IV and NASA/Goodyear MPP, typify this class. An MIMD (multiple-instruction multiple-data) computer consists of multiple processors (or fast time-sliced processing elements) driven by independent instruction streams and capable of branching independently but also able to pass data and synchronization signals between processors, perhaps via a shared memory. The Denelcor HEP, Cal Tech Homogeneous Hypercube, and numerous other machines belong to this class. SIMD machines are generally more efficient for applications characterized by very regular patterns of processing, in part because their "lockstep" mode of operation shares a single unit of control hardware among many data-processing elements and eliminates much of the message-passing overhead needed when all the processors of an MIMD machine must proceed through a computation in close synchrony. On the other hand, it is hard to make SIMD computers deal effectively with highly "branched" code, since the process of branching in such computers is normally imitated by temporarily disabling some of the processors in the parallel array, and thus several successive branches tend to reduce effective computational power drastically. MIMD processing arrays deal more robustly with complex branched code, since they are capable of executing not only innermost loops, but even complex branch-filled loops, in parallel.
2. *Visible versus "under the covers" physical communication pattern.* In all parallel machines some portion of the computation involves communication among the individual processing elements. From the physical point of view such communication is allowed in only a few specific patterns defined by the hardware. For example, the processors may be arranged in an obvious two-dimensional grid with each processor connected to its north, south, east, and west neighbor. A single operation on a machine wired in this pattern can only send a number from each processor to one of these neighbors. Proposed connection patterns for parallel machines include rings, n-dimensional cubes, binary trees, and various less obvious but much more powerful communication networks developed by communications researchers.

Once having chosen a communication net appropriate for the computational structure being developed, a parallel machine designer can choose either to make the structure of this net visible to the machine's user or to hide these details behind a set of programmed conventions that makes it appear that any processor can communicate directly with any other.

Both alternatives have arguable advantages. On the one hand, if a machine's actual communication pattern is well matched to an application and is made available to the programmer, it can be exploited to gain speed. Examples are the use of two-dimensional grid patterns for image processing and the use of various powerful communication patterns for optimal implementation of the Fast Fourier Transform. On the other hand, machines that relieve their programmer from direct concern for the pattern in which messages must flow

Copyright © National Academy of Sciences. All rights reserved.

between processors are easier to program for a wide range of problems, particularly problems characterized by highly variable patterns of communication. Another potential advantage of machines in this latter class is that the communication paths that they use, being hidden, can be varied dynamically, e.g., to bypass faulty components.

3. *Coarse-grained versus fine-grained designs.* In any parallel computer with multiple processing elements there is a cost tradeoff between the number and the size of the processors. The approach that stays closest to established single-stream processor technology is to use as few as possible of the largest available processors. The opposite approach is to achieve as much parallelism as possible by using a large number of small elementary processors, which can be single-chip microprocessors chosen to exploit the cost-effectiveness of a mass-produced item or, in the limit, can be a single-bit processor used in very large numbers. The decision to use a relatively small number of very-high-speed processors is justifiable by the argument that even largely parallelizable code will have significant serial sections, so high performance in serial code will be important for sustained overall high performance. Its polar opposite, the use of many single-bit processors, can be justified by a desire to push parallel execution to its limit, and by the surprising level of arithmetic power per square millimeter of silicon that one-bit processors attain. Perhaps the most important issue here is one of programming style. Since ordinary serial machines are coarse-grained, the technology for programming coarse-grained machines is understood best. Thus, it is plausible to expect a FORTRAN compiler to optimize code that keeps, say, 16 processing units busy, but not 16,000. On the other hand, if an algorithm is written with parallel processing in mind from the start, it may divide naturally into numerous processes, e.g., the processing of a 1,000 by 1,000 image in which relatively simple, low-precision algorithms must be applied to individual pixels or small pixel groups may fit most naturally on a million processor machine.

4. *Specialization of architecture to an anticipated application or programming style.* Scientific number crunchers are optimized for arithmetic operations on floating-point numbers, while processors intended for more varied symbolic applications must be optimized to handle irregular memory references and branching patterns, fast subroutine transitions, and fast context switching. They must provide run-time type checking and must support efficient execution monitoring during program debugging. Thus in designing a "scientific" processor one will emphasize certain types of hardware that are of little use in a "symbolic" machine, and vice versa. Various designs for large-scale "dataflow" computers, which are specialized in a different direction, illustrate this same general point. Dataflow programmming is organized around a set of fully parallel concepts. A program is regarded as an assemblage of "operators," all poised to fire as soon as they receive all their logical inputs. As soon as it fires, each such operator transmits its output to further operators, for which it becomes one of several possible inputs.

The innermost interpretation cycle for such a language consists of various operand-routing, instruction-dispatching, and instruction-execution activities, many of which can be overlapped. As compared to a general-purpose parallel assemblage, a cleverly designed layout of dataflow operation nodes in

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

memory and effective instruction dispatch hardware can probably double or triple the efficiency with which a dataflow language executes, while adding only minimally to the total hardware cost of a parallel system. These remarks typify the speed gains achievable by the architectural specialization of parallel machines.

V. SPECIAL-PURPOSE COMPUTING DEVICES

Extremely high computation rates can be attained efficiently by tailoring electronic hardware to algorithms of special importance (e.g., the Fast Fourier Transform) or to the requirements of a particular computer-intensive problem. Devices of this sort are of particular importance to diverse cutting-edge research efforts and to applications where dedicated equipment is essential but dedicated use of multimillion dollar general-purpose computers such as a large vector supercomputer would be prohibitively expensive.

These special-purpose computing devices point to a class of computing devices highly optimized to particular applications, which might be made possible by hypothetical future advances in compiling technology that allow automatic generation of whatever computing device is best suited for any specified application. Though this is far beyond present capabilities, manually designed special-purpose computing systems are of growing significance to defense uses, such as signal processing and image analysis, to high-energy-physics data acquisition and reduction, and to other areas.

It is steadily becoming easier to realize such systems, either by combining relatively standardized high-performance modules into specially wired configurations or by custom gate-array or VLSI designs. Special systems produced in this way tend to reflect end-user performance requirements closely, and hence they uncover important subproblems of the general supercomputer design problem quickly. Moreover, because of the emphasis on cost-effectiveness typically characterizing their design, they encourage industry to improve the effectiveness of more general computing engines supplied commercially. Also, though machines of this sort sometimes involve approaches that address only one specific problem, others affect larger classes of problems in an extremely cost-effective way, making them desirable attachments for general-purpose supercomputer systems. Encouraging a steady flow of such designs, which sometimes advance the effective frontier of computing in the application areas they address, can strengthen the U.S. competitive position significantly.

The sum of these special design efforts has become an important component of computer research that needs to be explicitly recognized and systematically cultivated. For this, an improved infrastructure for rapid and effective development of special-purpose systems is a prime necessity. In particular, projects motivated by special problems need a basic kit of standard modules out of which optimized architectures can be built. This should include low-cost single-board computers, high-density multiport memories, and high-speed switching modules that allow large communications networks tailored to specific applications to be fabricated rapidly. Some of these items are beginning to appear commercially, but further encouragement to industry and systematic attention to standardization issues are appropriate.

To increase the speed with which prototype special-purpose computational designs can be developed for significant experiments, system-level tools of the same quality as the best current VLSI design tools need to be developed. A focused program to develop such tools and make them available to universities, laboratories, and private-sector organizations appears to be an effective way to strengthen the work of special-system designers. Such tools would also facilitate the transfer of successful prototypes into indus-

Copyright © National Academy of Sciences. All rights reserved.

trially supported products, which is as important for special high-performance computational designs as for general-purpose supercomputer development.

VI. THE ROLE OF COMMUNICATION TECHNOLOGY

Supercomputers communicate three ways: internally, locally (in the same building), and outside (to machines elsewhere in the world). Existing technology for internal connections uses highly parallel buses and is limited primarily by wire length. Existing nearby machine telecommunications technology runs at a maximum of about 200 Mbit/sec, which is faster than most individual computers can accept information. Current outside connections are limited by available nationwide channels to 3 Mbit/sec, and 50 Kbit/sec is more typical.

In the next 5-10 years it is unlikely that the basic internal communication technology will change much. However, it is possible that within 20 years optical connections or wafer-scale integrated circuit technology will have significantly changed the way computers communicate internally.

Local interconnections will be affected by optical technology much sooner. It is possible that within the next 5-10 years commercial optical local networks running at a gigabit/sec rate will become available. This rate will not be accessible to any individual machine but will permit many supercomputers to share a relatively inexpensive network and exchange data at their internal bus rates.

NBS, DARPA, and cooperating international standards organizations have made large-scale networks possible by specifying protocols and standards that promote interoperability of heterogeneous equipment from different manufacturers. However, nationwide network speeds are currently inadequate for the amount of data consumed by supercomputers. Over the next 5 years the NSF ScienceNet project should be in place with high-speed satellite connections linking the major research universities. This will be a significant help to providing access to supercomputers.

Finally, we note that communication security on these nets may become an important issue, and this area deserves careful and early attention.

VII. ALGORITHMIC AND SOFTWARE ISSUES

Effective parallel processing will require a fundamental understanding of computational models, algorithms, programming languages, optimization techniques, processor-memory-I/O relationships, and interfaces to existing computing systems and end users. Though research on parallel algorithms is expanding, the lack of large-scale parallel machines for experimentation orients current work toward the study of individual algorithms rather than on practical applications, systems questions, or large software structures. Moreover, systematic work on large algorithm libraries, such as would quickly develop around major parallel computing installations, is hardly possible in the present atmosphere of largely theoretical studies. Also, although operating systems for parallel machines raise many subtle problems, it is difficult to address these except by working with real parallel systems.

Parallel programming is much more complicated than sequential programming because it deals with simultaneously occurring events involving potentially complex (time-dependent) interactions, so that elusive and nonrepeatable bugs can occur. For this reason, the usability of large parallel machines will be bound up with the development of programming languages that facilitate the expression of parallelism and the development of sophisticated optimizing compilers that improve the execution efficiency of such languages and find implicit parallelism where it is not explicitly expressed.

Hence, although initial experience (e.g., work on the HEP at the Los Alamos Labora-

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

tory) indicates that modest extensions to conventional languages will allow important applications to run with high efficiency on certain types of parallel computers, many new programming language concepts need to be investigated as ways for easing the further exploitation of parallelism. These include languages with vector and set operations (e.g., APL), dataflow languages, applicative languages, languages such as parallel LISP and PROLOG that emphasize recursive branching and backtracking processes, and languages in which communicating sequential processes play a central role. It is important both to encourage theoretical study and development of all these classes of languages and also to ensure that they become available for extensive experimental use as soon as substantial parallel computers suitable for them are constructed.

The present state-of-the-art software to support high-speed parallel computation is also represented by vectorizing compilers and more general parallelizing compilers capable of finding certain opportunities for parallelism automatically by analyzing ostensibly serial code. Both Cray Research and Fujitsu Ltd. have produced such compilers for their vector supercomputers, as have CDC, and U.S. research projects at the University of Illinois and Rice University.

While very useful, these systems are not yet able to automatically translate every standard sequential FORTRAN program into a good parallel program. Indeed, this will remain beyond the capabilities of automatic systems in the foreseeable future because the best parallelization usually requires substantial comprehension of the underlying problem domain. However, these systems do provide invaluable aids to the programmer because they permit him to write parallel programs cleanly in a high-level language without cluttering the language with special-purpose constructs.

Present automatic code vectorization techniques need to be evolved to handle programming for parallel multiprocessor supercomputers and to apply to whatever new parallel languages appear most important. Many of today's vectorization techniques are directly applicable; others will need to be adapted substantially or new methods invented. Program development environments tailored to the special requirements of new parallel systems and languages also need to be developed, especially to aid in the debugging of parallel programs. Graphics systems making it easy to deal with the complex outputs that parallel supercomputers can generate will also be needed.

Complex software developments cannot proceed far in a purely "paper mode." For them to flourish, real and relatively large-scale experiments with functioning hardware are essential. Thus, to accelerate the very challenging software research outlined in the preceding paragraphs, it is essential to deploy or provide remote access to experimental parallel machines as soon as they become available. Unconfining access is essential to the major systems software research and development efforts that will be required to ensure effective use of these new machines.

VIII. EFFORTS ABROAD

U.S. supercomputer developments must be appreciated in the context of efforts abroad, primarily those of the powerful and fast-moving Japanese computer industry, and also relative to an accelerating European effort. Our panel's conclusions concerning the Japanese program are as follows. The largest Japanese vector supercomputers (Hitachi S-810/820 and Fujitsu VP-200) have attained parity with, but not surpassed, corresponding U.S. machines such as the CDC Cyber 205 and the CRAY XMP. Japan's stated national commitment to advance in this area, the great engineering strengths of Japanese industry, and the leading position that Japan has won in certain lines of high-speed semiconductor component production will make for a very closely run race in this area. Nonetheless, the United States can at least expect to maintain

Copyright © National Academy of Sciences. All rights reserved.

parity in vector supercomputer production, and moreover in various rapidly expanding U.S. supercomputer use-oriented educational and research programs, e.g., the newly announced NSF Supercomputer Access Program, will help preserve U.S. leadership in vector supercomputer applications development.

Competition in regard to highly parallel supercomputers of newer experimental designs favors the United States even more strongly. The United States can expect to retain world leadership in this area, provided that agency support focuses on the most robust new designs and that proper bridges are built between university-centered innovation and the organizational and manufacturing capabilities of the U.S. computer industry. Without such links, it is feared that competition in this area will find various relatively small U.S. firms, including some recent start-ups, forced to compete with the extremely impressive technological capabilities of such major Japanese firms as Fujitsu, Hitachi, and NEC. It is noted in this connection that some Japanese firms have evinced interest in exchanging technology and development support for the results of U.S. conceptual research.

Software construction is also an area in which Japanese strengths, through not yet widely recognized in the United States, are already impressive, as shown, for example, by the first-rate vectorizing compilers available with the new Hitachi and Fujitsu vector supercomputers. Though the technology involved derives almost wholly from U.S. work, a recent study of Japanese software engineering practice, done by the JTECH Panel on Computer Science under the sponsorship of the U.S. Department of Commerce, concludes that in the production of standard software systems Japan may already be ahead of the United States and may be getting further ahead. For example, the average software production per employee in the 2,000-employee Toshiba software factory (which produces complex distributed systems process control software for power plants, steel mills, flight guidance, and so forth) is reported at 2,000 lines of high-level code per month, as compared to a typical U.S. figure of 300 lines per month. Error rates of 0.3 errors per 1,000 lines are achieved, as compared to U.S. rates 10 times as high, which may provide Japanese vendors the competitive advantage of marketing software with a 10-year warranty. Much higher rates of software module re-use than characterize U.S. practice are attained. The cited JTECH Panel report ascribes these significant successes to the systematic application of disciplined coding techniques and to source code control methods, quite familiar in the United States but much less systematically applied, and also to the sense of responsibility for a product that pervades Japanese society and to the factory discipline that Japanese management is able to apply to software production. This Japanese prowess in an area far more central to the computer industry than supercomputer production can potentially erode U.S. systems sales significantly and calls for further prompt analysis and appropriate response.

In spite of the intense international attention and publicity that has been focused on Japan's "Fifth Generation" computer effort, it is not clear to our panel that the ambitious goals set for it are matched by any entirely persuasive technical idea of how they can be reached. Nevertheless, it must be understood that even if not fully successful the kind of ambitious long-range research represented by the Fifth Generation project will generate significant new engineering, architectural, and software strengths for Japan. It must also be understood that Japanese universities and firms have mounted many parallel-computer explorations smaller and less publicized than the Fifth Generation effort and that cumulatively these will position Japan to move forward rapidly with new parallel machine designs. It is quite possible that these new Japanese designs will emerge immediately after the current wave of U.S. effort has

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

determined what lines of development are most promising.

We note also that it is clearly in the U.S. interest to build up a much larger body of computer specialists familiar with Japanese work and able to read Japanese technical literature. As long as tens of thousands of Japanese researchers follow the U.S. literature diligently while only a handful of U.S. computer scientists can deal with Japanese, the main flow of technical information will clearly run from the United States to Japan.

The European "Esprit" program does not appear to confront the United States with competition as severe as that seen in Japan. A review of the long list of projects to be supported under this program suggests that it will serve to strengthen European work in a broad range of technological areas, but concentrated work apt to result in supercomputer developments that will create major competition for the United States does not appear likely. Nevertheless, this European work needs to be followed closely.

The British response to the Fifth Generation challenge is the "Alvey" program, which is similar in its goals to the U.S. MCC research consortium now established in Austin, Tex. The four technologies making up Alvey are VLSI, software engineering, AI, and man-machine interfaces. Extensive university-industry cooperation is planned, and U.S. companies are welcome to participate as long as the research is done in Britain. Although the Alvey focus on supercomputer issues is not yet clear, it is important that this potentially significant national effort be followed closely.

IX. THE ARTIFICIAL INTELLIGENCE ISSUE

Construction of computerized artificial intelligences can justly be regarded as one of the most central long-term goals of computer science (and even of science generally), and it deserves tenaciously sustained support. The supercomputer developments now in progress will undoubtedly advance artificial intelligence research by putting new technical means at its disposal. An initially modest but growing degree of specialization in computer architectures can be expected to develop in response to the requirements of this field. Though the algorithmic structure of artificial intelligence computation is much less mature and stable than that of other areas of computation (e.g., numerical scientific computation or computer algebra), it is clear that work in artificial intelligence can be expected to emphasize random patterns of memory reference, various exact and approximate matching processes, computations with short quantities rather than multidigit floating-point quantities, and languages such as LISP whose execution can be accelerated (perhaps tripled) by suitable hardware. Moreover, the first stages of data input for subfields of artificial intelligence such as image and speech processing will require various special-purpose computational devices of the kind discussed in Section V, while exploration of various programming styles currently of interest to workers in artificial intelligence, e.g., fast interpretation of large systems of production rules, high-speed formula unification in logic, or neutral-net simulation, may also justify specialized hardware. The scientific computation and artificial intelligence communities favor distinct programming languages, emphasize different aspects of system performance, and prefer different operating systems, making it difficult for them to be served effectively by a single operational environment. Nevertheless, it remains difficult to derive any specific supercomputer designs from the requirements of artificial intelligence research, and in this sense the division of supercomputers into two distinct genuses of "scientific machines" and "artificial intelligence machines" is artificial. Perhaps it is best to say that artificial intelligence, as a young field in which approaches are still evolving, requires computers that can deal

Copyright © National Academy of Sciences. All rights reserved.

with rapidly shifting patterns of computation and communication, with great flexibility and at very high speed.

X. HOW TO MOVE FORWARD

Several major federal initiatives aimed at strengthening U.S. supercomputer-related capabilities have been launched and, if pursued along the strongest technical lines and with due regard for the importance of early and effective industrial involvement, can maintain U.S. computer preeminence. Principal among these are the DARPA Strategic Computing Program, the expanded supercomputer research programs of the Department of Energy, and NSF's Supercomputer Access Program. The steady support that NSF has provided for many, indeed most, of the university groups whose work has defined the principal supercomputer architectural alternatives now available has been of fundamental importance and is planned to continue and even to increase modestly.

Since DARPA's architecture program comprises the largest single federal project addressed to supercomputing, its activities will be broadest and most crucial. Work in many areas, including underlying technologies, improved design automation tools, fabrication means for both VLSI and broad-level special systems construction, conceptual exploration of new special general-purpose parallel computer systems, various small-scale prototyping efforts, and a few critical large-scale prototyping efforts, will all be sponsored.

Our panel is supportive of DARPA's intent to organize the selection of the culminating major efforts of this multiyear program via an orderly series of phases, beginning with conceptual studies aimed at the discovery of significant new computer architectures, progressing to the exploration of those architectures by simulation and experimental software development, then to small-scale hardware prototyping where justified, and finally, in the most promising cases, to the construction of large high-performance prototypes. We note that for these large culminating prototypes, research success, and even more the ability to move successful research results into broad U.S. use before foreign competition follows suit, will clearly be dependent on serious industrial involvement. The proper use of academic resources is in research and small-prototype development; the realization of large systems requires the skills of proven engineering organizations with established track records. For these reasons, we recommend that both DARPA and DOE encourage or require substantial industrial involvement (by either established or start-up companies) in the development of each experimental computer system chosen for large-scale prototyping. We understand that it is DARPA's intent to encourage this type of industrial involvement and we support this decision. Ideally, this should involve a substantial commitment of capital and crucial technical skills by the company (or companies) cooperating in such a development, to ensure that careful and independent private-sector technical judgment concerning project soundness is brought to bear in each such case. Moreover, to strengthen university-industry links and overcome what might otherwise be substantial familiarization delays in moving promising designs into large-scale prototyping, it is recommended that both DARPA and DOE encourage industrial involvement even in the initial smaller-scale stages of parallel computer prototyping. Early industrial involvement in special-purpose computer developments is also desirable.

Though considerably smaller than DARPA's effort, the DOE supercomputer program has maintained a significant focus in areas complementary to, and synergistic with, the DARPA effort. Historically, DOE has been the leading agency for the applications of large scientific computing and has maintained

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

particularly close relationships with the supercomputing community. Since optimization of computer architectures for floating-point computation and for the regular patterns of data manipulation characterizing many numerical applications may demand significantly different designs than are optimal for nonnumerical applications, it is important to maintain a strong program of architectural research and development with a clear scientific-computation focus. DOE's strong tradition and high level of expertise in scientific supercomputing make it appropriate for DOE to play a significant funding role for research on advanced computer architectures aimed at high-performance numerical computing, and acceleration of its program in this area is desirable. DOE is also in the process of significantly expanding its supercomputer access program to include scientists in High Energy and Nuclear Physics, Basic Energy Sciences, and Biological and Environmental Research. This expansion will contribute substantially to solving the national supercomputer access problem. In addition, DOE has acquired several commercial parallel computers and is making them available to many researchers. These DOE programs have played and will continue to play a very important role in maintaining overall supercomputer application capabilities.

As currently defined, NSF efforts will concentrate on making supercomputer capabilities broadly available to the U.S. scientific community, including but not restricted to computer scientists. Though initially this program will concentrate on providing access to existing vector supercomputers, our panel sees the NSF Access Program as having broad continuing implications and recommends that NSF plan to make all major experimental parallel computers nationally accessible as soon as feasible. Efforts of this sort, which serve to build supercomputer-use capabilities, strengthen the U.S. position by exploiting one of our most characteristic advantages, the great breadth and diversity of the U.S. computational community. However, the resources of this community will be heavily taxed by the deep algorithmic and software problems associated with the new classes of machines that will be developed over the next decade. Training of an expanded population of workers in this field must therefore be seen as an essential component of a balanced national effort.

The NSF Supercomputer Access Program is well calculated to serve the requirements of the large U.S. scientific community for whom computers are a tool for other research goals. However, in view of the great importance that NSF basic computer science support has had in the development of new architectural alternatives and the related computer science until now and in view of the excellent technical judgment with regard to hardware and software that NSF can bring to bear, the NSF programs should also be structured to support major software and applications development for the machines to which access is provided. Moreover, to avoid dissipating the attention of research groups among overly many new machines, all the agencies sponsoring supercomputer research and development should plan to support groups wishing to develop algorithms and software for experimental hardware systems produced commercially or by other universities. Such participating research groups should be guaranteed adequately "privileged" remote access, e.g., the right to take full control of a system for scheduled blocks of time to allow exploration of operating system changes in the light of new research ideas. In some cases, it will also be appropriate to disseminate smaller-scale versions of major experimental systems to several locations.

The need for quantitative measurements of progress and quantitative characterization of the advantages of alternative architectural approaches underlies all the areas of research we have discussed. It is therefore appropri-

Copyright © National Academy of Sciences. All rights reserved.

ate to encourage joint industry-university-government development of tools and techniques to measure progress in software, architecture, and algorithms for supercomputing systems.

XI. CONCLUSION

The U.S. national interest in the rapidly moving area of computer architecture should be as follows:

1. To ensure that all designs likely to succeed in any major way are explored actively. We must guard against technological surprise.
2. To get successful, or potentially successful, designs into the hands of large user and software development groups quickly, so as to accelerate the development of the necessary software. We must maintain system and software leadership.
3. To accelerate the transition of successful designs into commercial production, preferably by involving strong industrial groups in their development from the start.

It is quite important for U.S. activities aimed at maintaining awareness of Japanese technical developments through suitable translations be accelerated, possibly through programs undertaken by NSF or DARPA.

Finally, since the U.S. faces rapidly mounting competition in this area from both Japan and Europe, it is important that vigorously conceived programs, including those now planned by the major agencies involved, move forward without delay.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Report of the Research Briefing Panel on Information Technology in Precollege Education

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Research Briefing Panel on Information Technology in Precollege Education

John Seely Brown (*Co-chairman*), Xerox Palo Alto Research Center, Palo Alto, Calif.
James Greeno (*Co-chairman*), Graduate School of Education, University of California at Berkeley, Berkeley, Calif.
Henry Becker, Center for Social Organization of Schools, Johns Hopkins University, Baltimore, Md.
Richard Burton, Xerox Palo Alto Research Center, Palo Alto, Calif.
Allan Collins, Bolt, Beranek, and Newman Laboratories, Cambridge, Mass.
Andrea diSessa, Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, Mass.
*Harry Lasker, Interactive Training Systems, Cambridge, Mass.

*Dr. Lasker was unable to attend the meeting of the briefing panel, but met separately with the co-chairmen.

Additional advice was obtained from: Alan Lesgold, Lillian McDermott, Arthur Melmed, Andrew Molnar, Jack Palmer, Roy Pea, Senta Raizen, Paul Resta, Edward Silver, Charles Smith, John Tukey, and Fred Weingarten.

Mark Lepper, Department of Psychology, Stanford University, Stanford, Calif.
Milbrey McLaughlin, Graduate School of Education, Stanford University, Stanford, Calif.
Robert Parry, Department of Chemistry, University of Utah, Salt Lake City, Utah
Henry Pollak, Bell Communications Research, Inc., Summit, N.J.
Frederick Reif, Department of Physics and School of Education, University of California at Berkeley, Berkeley, Calif.

Staff

Norman Metzger, *Staff Officer*
Susan Newman, *Rapporteur*, The Learning Lab, WNET-TV Education Division
Gerry Kasarda, *Secretary*

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Report of the Research Briefing Panel on Information Technology in Precollege Education

SUMMARY

The development of more effective research methodologies and tools for the study of human learning, as well as recent advances in hardware capabilities, are opening up the possibility of using powerful new information technologies to aid individual learners and to help solve pervasive educational problems. Parallel work in artificial intelligence and the cognitive sciences has set the stage for qualitatively new applications of technology to education. What is required to move forward is increased support for basic interdisciplinary research, focused by the development of advanced learning systems employing the methodologies and equipment of artificial intelligence.

While success in this arena will not solve all educational problems, it will provide a new scientific basis for instructional and systems design, teacher training, and curricular restructuring. It also will create valuable new resources in the form of model electronic learning environments, while attracting a new cadre of professionals to the fields of education and educational research.

Thus, the panel recommends two federal initiatives to improve precollege education through the use of technology: (1) Support increased basic cognitive research on learning undertaken in conjunction with the development of advanced electronic learning environments employing the methodologies and equipment of artificial intelligence. (2) Continue federal support for individual scientists and applied research and development centers to analyze specific educational problems, develop and evaluate educational software, and find effective ways to integrate educational technologies into classrooms.

INTRODUCTION

Numerous recent reports have cited problems in the schools, declines in student performance on standardized tests, and the weakness of American mathematics and science education relative to other industrialized countries. Diverse sectors of society have begun to call for increased educational effectiveness and, in particular, for exploring the potential of information technologies for improving education.

However, earlier promises of educational improvements via technological interventions, such as educational television and computer-assisted instruction, have not

Copyright © National Academy of Sciences. All rights reserved.

always been fulfilled. Current investments in microcomputers are seen as limited in their effectiveness. Therefore, one charge of the present panel has been to assess what is different now—technologically, scientifically, and sociologically—that warrants increased federal attention to the research and development of instructional systems for the improvement of precollege education and what is required to advance the potential for learning represented by information technologies.

WHAT IS DIFFERENT?

While past projects to reform education based on learning research and technological advances have not always fulfilled the hopes of their designers, they have resulted in limited advances. Moreover, they have provided opportunities to learn how to achieve needed change more effectively. Recent studies of past reform efforts have identified specific difficulties surrounding school change and recommended how such difficulties can be surmounted. In addition, a confluence of several other factors is creating a unique opportunity to enhance the value of information technologies in precollege education.

Need for Educational Change

While observers are alarmed by evidence of declining "basic" mathematics and literacy skills among American students, there is simultaneously a growing emphasis on teaching "higher-order thinking skills"—reasoning, critical thinking, and learning about learning. These objectives require increased pedagogical emphasis on integrating basic skills with conceptual understanding, developing critical thinking and reading skills, enhancing the ability to see connections among objects and processes, and developing the capacity for self-directed learning.

Rapid Dissemination of Computers

The growing ubiquity of computers, while having at present relatively little direct educational impact, is nevertheless familiarizing a generation of teachers, parents, students, schools of education, school boards, and others with computers as tools and media. This sort of "computer literacy" sets the stage for the acceptance of more advanced educational applications such as those outlined in this report.

In addition, dramatic increases will occur in the computational power available to consumers at affordable prices within the next 5 to 10 years. These new levels of machine power make possible the delivery of the kinds of integrated and knowledge-based systems that are foreshadowed by present research and development efforts but that require concentrated basic research and experimentation for their full realization.

Effect of Technological Development on Research

Developments in the cognitive sciences and advances in computer technology have had a deep impact on educational research. Prior to 1960, learning research was concerned with primarily discovering the external conditions that influence readily observable behavior. While providing useful indications of the limits of human cognitive capacities and a scientific basis for the field of instructional design, behavioral psychology yielded little insight into the "black box": the underlying cognitive events that mediate behavior, including learning.

However, during the last 25 years, researchers have gained a new language and set of tools for expressing and validating learning theories that focus specifically on internal cognitive processes and their relation to the growth of knowledge and mental operations in humans. For example, symbolic computing languages used by the artificial

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

intelligence community, such as LISP, have provided a new notational system that is both scientifically precise and sufficiently flexible to represent complex learning processes.

In addition, computer technology has brought important changes in educational research by acting as a finely tuned "cognitive lab" for testing hypotheses and manipulating the large number of variables affecting human cognition. Through the simulation of human learning and problem solving, researchers have been able to verify detailed hypotheses about the mechanisms, architecture, and language of the human mind, particularly as they affect highly skilled performance by experts or the mental work by which novices gain new skills.

Finally, owing to the rapid growth of the information technology industry, a new generation of college and graduate students is becoming well versed in the computer sciences. The result is a larger pool of scientifically trained individuals from which, given adequate incentives, it is now possible to draw talent capable of conducting the combined educational and technological research needed to address new educational objectives.

GOALS AND OPPORTUNITIES

In addition to traditional competencies, students increasingly should learn the thinking skills to manage information, formulate effective probing questions, test hypotheses, make judgments, express themselves logically and lucidly, and solve problems.

Unfortunately, our present educational resources are inadequate to meet these objectives. In most classrooms, there is a lack of integration between low-level skills and higher-level understanding of a subject, and between disciplines such as mathematics and science, or writing and other subjects. Cognitive research confirms that knowledge and skills learned without conceptual understanding or functional application to problems are either forgotten or remain inert when needed in situations that differ from those in which they were acquired.

This common dissociation of formal schooling from experiential learning prevents students from mobilizing for formal learning the power of their intuitive understanding, even though they use it effectively outside the classroom. Students receive feedback in the form of grades, but are given little opportunity to revise their school work or to reflect on or communicate with other students or the teacher about their own writing, problem-solving, or other learning efforts. As a result, many students fail to correct even minor misunderstandings that nevertheless prevent continued advancement, and to develop skills for assessing their own learning problems.

While many of the needed changes could be effected by human teachers, appropriately trained, paid, and teaching smaller and fewer classes, the costs of this approach appear to be prohibitive. The possibility exists, however, that information technologies can provide the means to deliver some aspects of instruction not only cost-effectively but also in a deeply individualized and integrated way.

Specifically, parallel research in cognitive science and artificial intelligence is within reach of developing intelligent diagnostic, coaching, and discovery learning systems for a variety of subjects. Such systems could have a major impact on education. For example, knowledge-based simulation systems can enable students to test their understanding of scientific and mathematical principles, posing original problems that reflect their own questions about the world. With more powerful computers and research, it will be possible to create sophisticated work stations that can be used over a student's school lifetime. Such systems might include computerized lab-simulators, data-analysis software, writing and idea-structuring tools, and problem-analysis and management software.

In short, the real promise for the future is the development of integrated systems capable of interacting intelligently and sensitively with students and of providing a responsive,

Copyright © National Academy of Sciences. All rights reserved.

structured environment in which various skills and concepts can be developed and practiced. To realize this potential, initial investments must be made to capitalize research equipment and to support basic artificial intelligence and cognitive research focused on education and the development of educational technology. Such investments will enable us to move beyond the usual ad hoc applications of technology to educational problems. Besides providing a better basis for curriculum design, organization of subjects, and teacher training, increased scientific research will enable us to tailor instruction for individuals to a degree not generally found outside the apprenticeship learning models prevalent in graduate schools or other research centers.

TECHNOLOGY IN PRECOLLEGE EDUCATION

Information technology applied to education can be divided into three levels: systems of hardware and software now commonly used in schools; systems currently in the development and commercial pipelines; and advanced experimental systems embodying cognitive science and principles of artificial intelligence.

Level 1: Common Uses of Technology

Most of the computers presently available in schools are 8-bit microcomputers, with limited graphics, memory, and processing capabilities; however, some elementary and high schools are purchasing 16-bit and 32-bit "personal computers." While a growing range of applications is being developed, the majority of the schools that have microcomputers now use them to teach "computer literacy," BASIC programming, and for drill-and-practice or other forms of programmed instruction.

The best versions of the latter are based on behavioral analyses of the components of skill required to perform a given learning task, together with research on simple motivational factors and reinforcement schedules that influence persistence in a task. These types of programmed instruction are those most easily implemented on computers. In addition, these systems provide students with the advantages of sequenced practice, with the problems and text material adjusted to the level of student performance. Gains in student performance of routine skills, such as arithmetic calculation and elementary programming, have been accomplished through the use of these systems.

Assessment

While providing efficient means of acquiring rote skills, the educational benefits to be gained from these baseline resources are likely to be quite modest. The limited processing power and memory of microcomputers preclude their use for learning activities that conceptual understanding with mastery of procedural skills more effectively integrate; further, the analyses on which the programs are based do not address the cognitive mechanisms by which this sort of integration can be achieved. Computer literacy courses are rapidly outmoded as new hardware and software are developed. In addition, a new generation of more powerful and flexible programming languages and techniques is being created. As a result, instruction in BASIC may provide a poor model for students who might wish eventually to tackle complex programming problems.

Level 2: Recent Innovations

More imaginative educational uses of computers are being developed commercially or in cooperation with university-based researchers. These applications, which are being used in homes and experimentally in a few schools, include simple simulations, educational games aimed at reasoning and deductive skills, and simple word-processing programs. Some of them are based on cogni-

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

tive research that specifies learning problems that students have in a subject or the underlying skills that are needed for mastery of complex processes like writing. Others are based simply on the intuitions of skillful practitioners and instructional designers.

An example of a level 2 application is a simulation system, DynaTurtle, developed at the Massachusetts Institute of Technology. This system uses an extension of the programming language LOGO, in which students can give instructions that cause objects on the screen to move in accordance with the principles of Newtonian physics. Development of the system was based on the empirical recognition that students often fail to understand the fundamental concepts of a subject, even when they are able to apply formulas mechanically or to recite factual information. Cognitive research, some done using the DynaTurtle system, has documented the fact that it is often students' preconceptions about a subject that interfere with the acquisition of more accurate conceptual understandings, in spite of considerable formal instruction.

For example, students typically believe that an object in motion has some force that was transmitted to it by whatever caused the motion and that is dissipated as the motion continues. This experience-based "mental model" prevents most students from understanding momentum. Systems such as DynaTurtle take advantage of dynamic computer graphics to provide concrete information about general principles in response to student actions, enabling them to see the consequences of their ideas and to develop experience-based understandings of abstract concepts.

Assessment

Level 2 applications are being developed by educational researchers, often in collaboration with teachers and instructional designers, and frequently implemented by commercial organizations. While these instructional systems represent significant advances in educational software design, there are a number of ways in which they do not tap the full potential of information technologies to improve education:

- The systems often address isolated components of the curriculum and isolated instructional goals, and are thus difficult for teachers to integrate into classroom activities. Further, there are many learning topics for which an adequate cognitive analysis has not been performed yet, and little is known about how such specialized systems affect students' understanding of the instruction that is normally given on the same topics. Continued applied research is needed to address these concerns.
- Many of the systems have goals that are not thoroughly understood by teachers. Games addressing general cognitive skills are likely to puzzle those teachers who have a background in programmed instruction. A system for correcting a misconception in physics will seem mysterious to a teacher who shares the misconception, as many do.
- The systems that have been developed are based on ideas about helping students learn, rather than helping teachers teach. The orientation is a reasonable one, but is not optimal for encouraging enthusiastic adoption by teachers. Research and development focused on the use of computer-based materials in relation to the performance of teaching, either as a direct means of achieving teachers' goals or as supplementary materials that are viewed as beneficial, could increase the effective use of educational technology.

Level 2 applications represent a diversity of efforts that is important for discovering ways to use technology effectively for education. Besides their instructional value, such systems provide important artifacts around which to build continued research into the specific learning mechanisms that enable them to be effective. Most importantly, the involvement of teachers and instructional designers in this

Copyright © National Academy of Sciences. All rights reserved.

development makes it easier to begin implementing technological and other changes in the schools in an orderly way.

Level 3: Advanced Theory and Systems Development

Advanced systems combining symbolic computational capabilities, artificial intelligence, and cognitive theory have begun to be developed in a few well-equipped laboratories. As with level 2, each system addresses a relatively specific aspect of skill or knowledge but the systems differ in that they are informed by an explicit model of human learning and expertise. In addition to their eventual practical value as components of integrated learning systems, these systems and the computational equipment on which they are developed provide a powerful set of learning laboratories around which to deepen our knowledge of how conceptual change and understanding occur and of the methods by which we can bring them about more reliably.

Building on groundwork by behavioral scientists and interdisciplinary work in linguistics, psychology, and computer science, developments in cognitive science have coalesced around a theoretical model of the human mind as a complex information-processing system in which information is organized, stored, and operated on in ways that we are only now beginning to characterize in detail. Within this framework, learning may be viewed as a process of information transfer, leading to useful lines of inquiry concerning the structure of information to be transferred, the characteristics of the receiver (including the biases and constraints of the "system architecture"), and a theory of the transfer process itself.

At the same time, methods in artificial intelligence for constructing expert systems have developed rapidly. Two features of this work are important for education. First, expert systems in a given subject are useful components of computerized tutoring systems; their commercial development, therefore, represents a potentially valuable resource for education. More importantly, the construction of an expert system requires an explicit representation of the knowledge that constitutes expertise. This work, along with empirical studies of human expert performance, has been a means of gaining crucial insights into the nature of human knowledge, skilled problemsolving, and reasoning.

Expert Tutoring Systems

Early efforts to adapt expert systems to education foundered because the organization of factual and procedural knowledge within the systems bore little resemblance to human thinking and reasoning. Further work informed by cognitive research has succeeded in specifying more completely the types of knowledge required for mastery of problemsolving in a given subject, and the results have been applied to education and training. For example, developers of NEOMYCIN refined, restructured, and semantically rationalized the knowledge base of the seminal expert system MYCIN to teach medical diagnosis to students.

The panel believes that continued efforts to formalize the conceptual knowledge, problemsolving strategies, and inferential processes of experts will enable the creation of more robust systems capable of solving novel problems in a given topic or domain. Such systems should be able to tackle unanticipated, student-posed problems and, perhaps more importantly, to show students the underlying processes by which the systems choose among alternative solutions.

Expert Diagnostic Systems

Detailed specifications of the knowledge and skills required for work in a discipline make possible the construction of expert diagnostic systems that are able to isolate

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

underlying causes of student errors. Construction of such systems requires a complete theory and symbolic representation of all the components of an overall skill. However, it also requires (a) a theory of how those subskills are mislearned that accounts for observed student mistakes; (b) a theory accounting for student modification—e.g., shortcuts—of formally taught procedures; and (c) a model of the "noise" in the students' execution of the skill.

One such system is DEBUGGY, an expert system for diagnosing procedural errors or "bugs" in base-10 subtraction. Starting from its data base of more than 130 empirically determined "primitive" or single bugs, DEBUGGY synthesizes a procedural model of the exact nature and cause of a student's deviation from a correct subtraction procedure. What makes the development of such a system especially noteworthy for education is its robustness: its ability to handle compounded student errors that arise when students have more than one bug, as well as their careless mistakes that can disguise systematic errors. Tested with over 4,000 students, DEBUGGY is able, using a set of search heuristics, to distinguish among 10^8 possible hypotheses about the nature of a student's errors.

Besides possible direct use in the classroom, computer-based diagnosis is an important part of developing cognitively based testing methods that measure a student's partial knowledge and pinpoint his misconceptions. The information provided by DEBUGGY can be used to provide personalized and precise remediation not hitherto possible.

Such systems also have been adapted for use in training student teachers, focusing their attention on the underlying causes of student learning problems and on strategies that they themselves can use in diagnosing students.

Perhaps most importantly, diagnostic systems provide a new tool for validating precise theories of learning. Previously unobserved knowledge states can now be deduced, enabling learning theorists to postulate changes in mental states and to verify them on the basis of successful prediction about observable behavior.

Expert Coaching Systems

The successful construction of expert systems and diagnostic systems in various domains has enabled the construction of computer-based coaches. Existing computer-based coaches such as WEST and WUMPUS construct a differential model of the student's conceptual strengths and weaknesses both through error diagnosis and by comparing student behavior with what the computerized domain expert would do at any given point in solving a problem. However, effective coaching also requires knowledge about how and when to intervene. Although such knowledge is usually held tacitly, even by master teachers, existing computer-based coaching systems have begun to characterize this kind of knowledge as a collection of rules that govern computer intervention.

Future work will be directed toward meeting two major challenges. The first one is to expand the current existence proofs for expert coaching systems to include a wider variety of domains, more flexible and psychologically informed tutoring rules, and more robust domain expert systems that mimic closely the reasoning processes of human experts. Such systems will provide a computational workbench for refining our understanding of tutoring knowledge and a building block for commercial educational software developers.

Secondly, systems must be both cost-effective to construct and readily modifiable by end-users. The emergence in artificial intelligence of new programming tools makes it possible to reduce the production time for a domain-specific coach from years to months. In addition, use of these sophisticated programming environments can enable researchers to make the system's tutoring

Copyright © National Academy of Sciences. All rights reserved.

knowledge more accessible and modifiable by teachers and educational technologists.

Discovery Learning Environments

Systems are also under development employing artificial intelligence techniques and based on cognitive research that does not require direct computer intervention into the learning process. Instead, these systems provide a rich environment that aids students in discovering how to solve problems more effectively by recording, structuring, and "playing back" the student's own problem solving processes.

Such systems are based on cognitive research that shows that successful learners exhibit a high degree of reflective skill, reviewing their progress toward a goal, analyzing their misunderstandings, and revising their strategies accordingly. Poorer learners tend simply to skip over or invent temporary solutions to aspects of a problem that they do not understand.

Computational systems have been developed in algebra and geometry that handle low-level computational operations, thus enabling students to focus attention on their strategic choice as they solve problems. The systems then provide a structured "trace" of their solutions so that students can see the alternative paths that they have tried. Preliminary tests with students show that they are motivated to review their choices and to formulate hypotheses leading to more effective choices in later problems.

The capability of information technologies to record and abstract students' problem-solving processes provides unique leverage for researchers interested in the role of error in learning, while providing students with a mental "mirror" of their own thinking. Armed with these tools, researchers can determine empirically the effects of increased self-reflection and of extended work and revision on fewer problems—particularly for poorer learners.

Such learning environments are impractical with the technology currently used in schools. But technology is already available in research laboratories that could make apprenticeship and experiential learning possible on personal computers of the late 1980s, along with methodologies to develop the scientific understanding of learning needed to design the computational environments and to formulate effective learning tasks.

The technology will be here; the need is for an adequate scientific understanding of how to exploit it. The next section describes five specific research problems on which the field is ready to make significant advances.

RESEARCH NEEDS

Conceptual Change

Current educational practice appears to be remarkably unsuccessful at connecting school knowledge with experiential knowledge. A major research task is to redesign the educational materials, taking into account the deeply held beliefs and conceptions that students have when they begin instruction. Characterizations, already begun, of students' naive conceptions are not enough to enable remediation. A more adequate understanding of how change in the organization and representation of knowledge occurs during learning is needed.

Students' initial conceptual framework also includes powerful conceptual categories, called "phenomenological primitives," by which they organize their understanding and learning. Growing understanding of the function and origins of these structures can improve our ability to craft educational experiences and restructure domain knowledge so that students can integrate newly acquired knowledge into the informal knowledge and reasoning that they already have, while preparing for the deeper conceptual changes necessary for correct understanding.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Changes in Subject Matter Content

Widespread use of computation is leading to substantial changes in both the form and the content of procedural skills that it is important to teach to students. Obvious examples arise in mathematics, where procedures for calculation that are difficult to learn can be performed routinely on inexpensive calculators. Somewhat less obviously, many procedures that are taught in mathematics are done to facilitate hand calculations. For example, many procedures for "simplification" in elementary algebra produce expressions that are simpler if one has to evaluate them on paper, but are by no means simpler conceptually or for operations with a digital computer.

In addition, a major reexamination of the curriculum is called for as a result of new opportunities for integrating the content of education both within and between subject-matter disciplines. For example, the capabilities exist for creating educational data bases that can manifest connections between parts of the curriculum. Such connections are rarely made explicit. For example, Ohm's law is usually taught in elementary physics without reference to its being an example of a simple proportional function in mathematics. Design of curriculum sequences with relations like this in mind could increase significantly the coherence of the educational experience and the knowledge that results from it. Cross-referenced data bases also would aid teachers, who could generate easily examples of concepts in their instruction that relate to other areas of their students' work.

Measurement and Evaluation

A critical goal of research on learning is the development and testing of methods for measuring deep cognitive outcomes to guide the design of instructional systems. Current diagnostic techniques enable us to isolate misunderstandings in the acquisition of procedural skills at a new level of detail; however, the way in which more general misunderstandings develop and how they affect subsequent learning is less well understood. Research is needed to assess students' grasp of the conceptual bases and purposes of these procedural skills that they learn to perform, and to develop tests of their ability to modify a procedure flexibly when required.

The ability of sophisticated computational systems to record and abstract the actions of a student while engaging in various complex activities provides a rich source of data for learning researchers. The analysis of verbal protocols—that is, stream-of-consciousness records of subjects' thinking while performing complex cognitive tasks—has led to many advances in cognitive science. With information technologies enabling the automatic collection of electronic protocols and providing inferencing tools to aid in analyzing student thought processes, we are likely to achieve a new set of advances.

Motivation

In addition to cognitive factors, intrinsic motivation is important to successful learning. One counter-intuitive result of recent research indicates that extrinsic reward—for example, an extra play period—can reduce the motivation of students who already engage in a particular activity willingly.

While motivation theory has traditionally been a "messy" and difficult area of scientific inquiry, the use of electronic learning environments as motivational laboratories, in which variables and effects can be isolated and manipulated subtly, has had a significant impact on research already. For example, study of the motivational factors of computer games—level of challenge and user control, completeness of information and its effect on curiosity, and the role of fantasy—establishes a beginning taxonomy of variables that can be examined within the context of exper-

Copyright © National Academy of Sciences. All rights reserved.

imentally designed learning environments. The popularity of fast-action video games and their possible consequences for the cognitive development of American youth necessitates a better understanding of how the slower, more difficult mastery of cognitive skills can be made rewarding for a broad range of students.

Research is needed also to assess the implications of what appear to be the motivating effects of the computer as a medium, independent of instructional content, and to determine its impact on the transfer of motivation and learning to nonelectronic settings. Work also is needed to determine the interactions between factors such as aptitude and a given motivational treatment within the electronic learning context and to experiment with feedback and reward structures.

Implementation Issues

As stated earlier, many past efforts to reform education have had only limited success. Analyses of educational change have identified a set of basic pedagogical techniques and curricular contents that have been particularly resistant to change. Cognitive analysis suggests that, in many cases, these are precisely what must change if improvements in the quality of learning for all students are to take place. How to achieve these changes is therefore a crucial research issue.

Appropriate sorts of demonstration projects can help to spread awareness among educators about the types of learning activities, motivational effects, and other qualitative outcomes that are possible with flexible electronic learning systems. Work at centers focusing on the use of educational technology in schools also should contribute to our knowledge about how to effect institutional change.

Curricular and pedagogical change in the direction of individualized discovery learning and greater integration among subjects are impeded by a lack of objective measures of the benefits of those forms of instruction. With the development of validatable measurements for deep cognitive outcomes and demonstrations through testing of the positive effects of cognitively based electronic learning, it may be possible to generate a discerning market demand for fundamentally new kinds of learning environments. Such an approach to needed reform has the clear advantage of mobilizing market forces, creating an accountability structure for the educational software industry, and preserving local choice in formulating school policy.

New learning systems themselves can become the carriers of cognitive theory and new principles of pedagogy into classrooms and homes. Teachers will then be able to assume a freer role in aiding and assisting students in their individual efforts. Further, teachers can use technology to improve their teaching techniques. Cognitively based learning systems can be used simultaneously to teach professional courses in education, to engage teachers in the same kinds of activity-based learning experiences that they will use with students later, and as objects of study about underlying cognitive processes.

Information technologies can have a profound impact on the sociology of the classroom. Several pilot studies have shown that the use of computers as tools and as communications devices tends to alter relationships between students and teachers and among peers. The design of collaborative activities around learning systems can help engineer a new classroom infrastructure within which innovations are less disruptive.

CONCLUSIONS

The time is ripe to develop a new scientific discipline as a fundamentally new resource for American education. A multiyear effort focused on advanced basic research using information technology to enhance and study learning could provide a major synthesis of the knowledge, theory, and methodology that now exist in the relatively separate disci-

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

plines of artificial intelligence, cognitive psychology, linguistics, and educational research in school subjects. Integral to this research synthesis will be experimentation with and construction of advanced computer-based learning environments. Such environments will serve as laboratories for deep cognitive research, while also providing significantly new educational resources directed at real educational problems.

By developing new learning environments and evaluating their pedagogical effectiveness, gains will be made in understanding the mechanisms of experiential learning, the integration of formal and informal knowledge, the integration of conceptual knowledge and procedural skill in a domain, and the integration of knowledge across present subject matter disciplines. The technological and theoretical products of this work will provide models of learning systems that can increase the number of American students who succeed in acquiring the knowledge and cognitive skills necessary for effective work and citizenship.

Resources Required

In order to make these advances, increased resources of three kinds are needed, beyond the levels that are currently available in the field:

1. *Sophisticated Symbolic Computation*

Wider access to sophisticated computational equipment of the kind now used in standard artificial intelligence research is required for effective contributions to the research and development effort. At present, the availability of such equipment to cognitive and educational researchers is extremely limited, and well beyond funding allocations in these areas. The computational resources needed for cost-effective research and development exceed those required for the delivery of educational systems in schools, if they are to support the interactive programming methodology that makes possible rapid exploratory and experimental software development. Therefore, we recommend a substantial increment of funding to make sophisticated interactive computation and associated networks much more widely available than is now the case.

2. *Interdisciplinary Research*

Research on learning and on the development of educational systems development requires the talents of computer scientists, cognitive scientists, artificial intelligence researchers, social scientists, and subject matter experts. At present, opportunities for these sorts of interdisciplinary partnerships are limited and occur primarily as isolated projects in universities. A systematic effort to establish an integrated discipline, composed of individuals trained in the relatively new theoretical and technical skills needed to contribute effectively to the growth of knowledge about learning and its applications, is necessary. The creation of one or two special research centers (perhaps modeled after the Institute for Theoretical Physics at Santa Barbara—and similarly established with support by the National Science Foundation or other federal agencies) would provide powerful interdisciplinary resources to further the timely growth of the scientific knowledge requisite to significant educational applications of advanced technology.

3. *Stable Funding and Talent*

Existing programs represent a useful range of perspectives for near-term research and development, and it seems appropriate to continue with that mixture of support. However, current funding levels reflect a very low priority for basic research. Recent oscillations in funding levels at the National Science Foundation (Science Education Directorate) and at the National Institute of Education have been catastrophic. Stable funding for basic research and training is essential if this field is to attract first-rate scientists and graduate students. Programs are needed to support postdoctoral fellowships in which individuals trained in one of the relevant disciplines could visit for extended periods research

Copyright © National Academy of Sciences. All rights reserved.

institutions where they could obtain training and experience in the other fields.

General Comments

The levels of computer hardware and instructional programs now prevalent in schools, while probably beneficial, do not add sufficient value to ordinary instruction to warrant increased federal investment to promote the distribution. The more fruitful use of federal funds for distribution of technology is to continue to encourage development and integration into schools of level 2 and level 3 systems. The existing programs are needed to identify specific instructional difficulties and opportunities for improvement and to increase the use of instructional systems in schools and other educational settings.

The advanced research and development efforts suggested here are intended to complement programs that are already in progress or planned at federal agencies. Certainly, existing efforts will be enhanced significantly by the increased understanding of learning, epistemology, and advanced symbolic computation that would result from more advanced projects.

We are, therefore, suggesting a clear recognition of (1) the need to capitalize sophisticated symbolic computation equipment for use both as research tools and as hosts for experimental learning environments; (2) the interdisciplinary nature of basic research in this area; and (3) the need to attract first-rate scientists and graduate students to the field. Significant increments in research capabilities can be achieved with relatively modest amounts of additional funding for equipment and to support interdisciplinary research and postdoctoral fellowships. The results could bring about major advances in our ability to use information technologies in precollege education.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Report of the Research Briefing Panel on Chemical and Process Engineering for Biotechnology

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Research Briefing Panel on Chemical and Process Engineering for Biotechnology

Arthur E. Humphrey (*Chairman*), Lehigh University, Bethlehem, Pa.
James E. Bailey, California Institute of Technology, Pasadena, Calif.
Harvey W. Blanch, University of California at Berkeley, Berkeley, Calif.
Stuart E. Builder, Genentech, Inc., San Francisco, Calif.
Robert R. Dorsch, E. I. du Pont de Nemours & Company, Inc., Wilmington, Del.
Wolf Hanisch, Cetus Corporation, Emeryville, Calif.
James Lago, Merck & Company, Inc., Rahway, N.J.
Alan S. Michaels, Alan Sherman Michaels, Sc.D., Inc., New York, N.Y.
John A. Quinn, University of Pennsylvania, Philadelphia, Pa.
Kevin Ulmer, Genex Corporation, Rockville, Md.
Daniel I. C. Wang, Massachusetts Institute of Technology, Cambridge, Mass.
Neal R. Amundson (*Ex-Officio*), University of Houston, Houston, Tex.

Staff

Robert M. Simon, *Project Director*
Robert M. Joyce, *Rapporteur*
Robin Illsley, *Administrative Secretary*

Copyright © National Academy of Sciences. All rights reserved.

Report of the Research Briefing Panel on Chemical and Process Engineering for Biotechnology

EXECUTIVE SUMMARY

The phenomenal progress in molecular biology, genetics, and biochemistry over the past two decades now makes possible the programming of living cells of virtually every type—microbial, plant, and animal—to generate products ranging from simple molecules to complex proteins. The United States leads the world in research in this "new" biology. Maintaining that leadership and commercializing the fruits of that research are vitally important to the future economic well-being of the nation and to its international technological stature. A tremendous number of opportunities await commercial exploitation. This will not occur rapidly without adequate knowledge of the fundamental engineering principles underlying bioprocess scale-up. What is needed is a knowledge base in process engineering that combines the skills of the biologist and the chemical engineer in order to bring these opportunities to fruition. The rapid establishment of this knowledge base will require the cooperation of academia, industry, and government.

The intense international competition in commercializing high-technology opportunities, particularly in biotechnology, highlights the need for the United States to maintain its scientific and industrial leadership. Despite this nation's preeminent position in basic biological science, our foreign competitors have established commercial positions by practicing effective and forward-looking biochemical engineering. They can be expected to be no less aggressive as the results of U.S. research in the "new" biology diffuse abroad.

Estimates of the potential worldwide markets for biologically derived products fall in the range $40-$100 billion annually by the year 2000. First entry into these markets will be critically important, and major shares of the worldwide bioproducts market will be captured by those countries who possess the needed national capability in engineering research and personnel.

There are not enough biochemical engineers in the United States today, nor are there enough faculty to train the engineers who will be needed by the emerging biotechnology industry. Fewer than 20 U.S. Departments of Chemical Engineering have meaningful biochemical engineering programs; collectively, these are training fewer than 60 Ph.D. and M.S. graduates annually. This panel estimates that the annual need for graduate-

Copyright © National Academy of Sciences. All rights reserved.

level biochemical engineers over the next decade will average 2-3 times that number. Moreover, the training of biochemical engineers should expose them to the power of modern biological science in manipulating and controlling cellular biosynthesis, and to the utility of life sciences techniques and methods in solving critical large-scale bioprocess problems. Prompt and effective exploitation of the "new" biology is critically dependent on the improvement of this life science/ biochemical engineering interface.

The technology needed to capitalize on the discoveries in modern biology falls in two areas:

- Increasing our biochemical engineering expertise in the design, scale-up, and optimal control of large-scale processes and culture of bacterial, plant, and animal cells.
- Expanding our knowledge base so that we will be able to achieve the large-scale recovery and purification of complex, unstable biological macromolecules, as well as of simple biologically derived organic compounds (e.g., ethanol, amino acids) from the dilute and impure solutions that are the initial products of biosynthesis.

Our recommendations for government action are aimed at developing this needed knowledge base in biotechnology and at reducing the communications gap between life scientists and biochemical engineers. Specifically, we recommend that appropriate federal agencies:

- Provide funding at adequate levels for basic research in biochemical and process engineering.
- Establish grants for cooperative, cross-disciplinary research that will stimulate the development of research groups led by joint principal investigators from both the life sciences and biochemical engineering.
- Provide funds for academic units to acquire, operate, and maintain the special types of equipment needed for biochemical engineering research.
- Establish faculty development awards to allow tenured faculty to broaden their expertise in biotechnology through special study and leaves, and to provide promising young faculty with the opportunity to immerse themselves in biotechnology research by temporarily relieving them of some of their teaching commitments.
- Establish institutional grants to train graduate students, thus providing the personnel that will be needed if U.S. industry is to capitalize on emerging opportunities in biotechnology.

INTRODUCTION

The phenomenal progress in molecular biology, genetics, and biochemistry over the past two decades now makes possible the programming of living cells of virtually every type—microbial, plant, and animal—to generate products ranging from simple molecules to complex proteins. The United States leads the world in research in this "new" biology. Maintaining that leadership and commercializing the fruits of that research are vitally important to the future economic well-being of the nation and to its international technological stature. The utilization of living cells and their enzymes for the industrial manufacture of useful products—for which the term "biotechnology" has been coined—requires the adaptation of these systems to large-scale production facilities. This transformation from laboratory experimentation to commercial production is the province of the chemical engineer, or more particularly the biochemical engineer.

Biochemical engineering translates the knowledge developed in biological and chemical science into practical industrial processing; the biochemical engineer bears the same relationship to the life scientist that the chemical process engineer does to the chemist. Chemical engineering enriched by active involvement in the "new" biology can provide a knowledge base to meet the challenges that will be faced by the next generation of chem-

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

ical engineers in commercializing biotechnology. The establishment of this knowledge base will require the cooperation of academia, government, and industry. Chemistry and chemical engineering have enjoyed a long tradition of active cooperation to the benefit of the chemical industry. However, the interface between chemical engineering and biology needed for practical implementation of biotechnology has not yet evolved. What is needed is a sound curriculum that will promote this cooperation and train the biochemical engineers who can develop the knowledge base required to serve the biotechnology industry.

OPPORTUNITY ASSESSMENT

The opportunities for commercialization of the products emerging from the "new" biology are both tantalizing and highly diverse; among the more exciting prospects are applications falling into the following five categories:

Human and Animal Health Care. A revolutionary new family of diagnostic products based on enzymes, monoclonal antibodies, and other genetically engineered proteins promises to provide quick and highly accurate detection of immunity to or infection by viral and bacterial diseases, of susceptibility to autoimmune diseases, of the presence of genetic defects, or of the existence of neoplasms. Other significant opportunities include novel prophylactic products, exemplified by vaccines for the prevention of viral, bacterial, and protozoal diseases such as hepatitis, typhus, and malaria; new therapeutic biologicals for the treatment of cardiovascular and cerebrovascular disease, neurological and CNS diseases, rheumatoid arthritis, diabetes, and cancer; peptide hormonal substances that increase milk production of dairy cattle and stimulate growth, fecundity, and enhanced feed utilization of cattle and other farm animals.

Human and Animal Nutrition. Applications include the utilization of low-cost carbon sources for new microbiological and enzymatic syntheses of amino acids, sugars, and edible fats and oils, and the use of large-scale fermentation of low-value feedstocks in the manufacture of nutritionally balanced single-cell protein for human or animal consumption.

Agricultural Chemicals and Related Products. Prospects for applications of biotechnology include biologically derived fungicides and herbicides that are highly potent, highly specific, and environmentally safe; plant growth regulators to stimulate crop productivity, reduce plant sensitivity to environmental stress, and reduce fertilizer requirements; new techniques for crop propagation and strain improvement that employ genetically manipulated plant cell clones to replace seed.

Environmental Improvement and Protection. New microbial and enzymatic techniques for removing or destroying toxic pollutants in municipal and industrial wastes provide a promising horizon for the application of biotechnology to environmental problems.

Natural Resource Utilization. New microbiological processes for the recovery of metals and nonmetals (e.g., iodine) from low-grade ores and subsurface aquifers are under development, promising to increase substantially our supply of essential minerals. The possibility of increasing recovery of petroleum from depleted reservoirs by microbiological means is being explored, as are *in situ* microbial and enzymatic techniques for transforming solid fossil carbon sources (coal, shale) into gaseous and liquid fuels.

These tremendous opportunities cannot be exploited without adequate knowledge of the fundamental engineering principles underlying bioprocess scale-up. We face critical engineering problems in the design and control of large-scale cell-culture processes for the manufacture of new bioproducts, and of processes for the efficient large-scale recov-

Copyright © National Academy of Sciences. All rights reserved.

ery and purification of bioproducts from the complex mixtures in which they are made.

INTERNATIONAL COMPETITION IN RESEARCH

Who will lead the commercialization of the "new" biology? U.S. progress in basic biochemical engineering research has lagged substantially behind its progress in the life sciences. The United States has also lagged behind Western Europe and Japan in support for basic chemical engineering research. This situation must be corrected if our national ability to lead the commercialization of the "new" biology is not to be seriously compromised. Otherwise, the Japanese and the Europeans will certainly take the leading role. The following case histories are illustrative.

- The U.S. consumer was spared much of the world sugar crisis through bioreactor technology based on Japanese patents. This technology uses a continuous, immobilized-enzyme process that converts the glucose derived from domestic cornstarch to high-fructose corn syrup, a commercial sweetener. The process, with $1.5 billion annual sales, now uses $50 million annually of immobilized enzyme purchased from European companies.
- U.S. companies led the world in manufacture of amino acids prior to the 1960s. Then, an intensive Japanese effort began to improve amino acid production through biological and biochemical engineering techniques. First, classical genetic strain improvement techniques were applied to isolate mutant microorganisms that produce greatly increased quantities of amino acids. Second, biochemical engineers utilized very large bioreactors and developed efficient methods for recovering the product. Today, Japan dominates a business with annual world sales of $1.7 billion.
- Process technology for penicillin manufacture was developed in the United States in the 1940s. Protected by patents, several U.S. companies dominated the world penicillin market for many years. Now that the patents have expired and U.S. companies have decided to invest in research on new, higher-return products rather than on process improvement, European companies with modern process technology account for the preponderance of worldwide sales of natural penicillin.

Despite the preeminent U.S. position in basic biological science, our foreign competitors have established commercial positions by practicing effective and forward-looking biochemical engineering. They can be expected to be no less aggressive as the results of U.S. research in the "new" biology diffuse abroad. Foreign developments in biochemical engineering already attest to that aggressiveness.

- Basic technology for membrane separation of biomolecules was invented in the United States, but the Germans and Japanese are the ones applying it to separations of enzymes and amino acids from complex mixtures.
- Technology for very large (400,000 gal) continuous fermenters was developed and is being practiced in the United Kingdom. This development pushes biochemical engineering to limits not yet explored in the United States.
- Although the use of fermentation to produce ethanol is an ancient technology, more efficient immobilized-cell, continuous processes have been conceived, and Japan has established the first demonstration-scale plant.

A survey of biochemical engineering research and education in the United States, Western Europe, and Japan was carried out by several members of this panel. A striking difference between U.S. and overseas activities is the magnitude and organizational structure of the government-supported research and development effort in biotechnology. West Germany, Japan, and the United Kingdom each have three federally supported institutes dedicated to biotechnology; there are none in the United States. These nine

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

institutes bring together academic and industrial investigators, feature cross-disciplinary activities, and have impressive operating budgets.

In 1983, the West German Gesellschaft für Biotechnologische Forschung (GBF) had an operating budget of DM37.1 million ($14 million), and the Institut für Biotechnologie II in Jülich had DM11.1 million ($4.3 million) for biochemical engineering alone. The last figure nearly equals the entire NSF annual budget in biochemical engineering. A quantitative comparison of space and associated equipment for biochemical engineering shows that West Germany's research space is nearly double that of the United States.

In Japan, the MITI 10-year plans have strong research and development components in biochemical engineering; three of the nine areas emphasized are bioreactors, animal cell culture, and membrane separation. Japanese government support of membrane separation research and development alone amounted to $20 million in 1983. The comparable NSF effort is less than $1 million.

It is obvious that countries such as Germany and Japan are laying a foundation of research and trained personnel as part of their strategy for intense international competition in the arena of biotechnology. The potential economic rewards for success are very great; estimates of the potential worldwide markets for biologically derived products fall in the range $40-$100 billion annually by the year 2000, corresponding to about 15 percent of the estimated total annual market for chemicals. First entry into these markets will be critically important in international competition, and major shares of the worldwide bioproducts market will be captured by those countries who possess the needed national capability in research and personnel.

STATUS OF U.S. RESEARCH AND EDUCATION

There are not enough technologically competent biochemical engineers in the U.S. today, nor are there enough faculty to train the engineers who will be needed by this emerging industry. According to an analysis by members of this panel, there are roughly 80 chemical engineering faculty members in the United States who currently have some involvement in biochemical engineering research and at least 50 more who have indicated an active interest in the field. Fewer than 20 U.S. Departments of Chemical Engineering have meaningful biochemical engineering programs; collectively, these are training fewer than 60 Ph.D. and M.S. graduates annually. Yet this panel estimates that the annual need for graduate-level biochemical engineers over the next decade will average 2-3 times that number. The current personnel demands of industry are recruiting younger faculty members and recent graduates who would otherwise be strengthening the university teaching and research capabilities in this rapidly growing field.

The costs of doing research in some areas of biochemical engineering are increasing rapidly as the field moves away from an Edisonian approach to an emphasis on discovering and applying fundamental engineering principles. Biochemical engineers are beginning to investigate the principles underlying the design and control of reactors for growing mammalian cell cultures. Such work holds great promise in reducing to practice the substantial U.S. lead in monoclonal antibody research and in large-scale cell culture for the production of tissue plasminogen activator for the rescue of heart attack victims. Virtually no university in the United States has cell culture laboratories equipped for engineering studies. Such laboratories are costly. A critical review of what constitutes adequate funding to perform state-of-the-art research should be part of an overall plan for ensuring U.S. competitiveness in this area.

Most important, there is a critical need to strengthen the life science/biochemical engineering interface. The historic success of the development of the U.S. chemical industry has in large measure been due to close collaboration and communication between the

Copyright © National Academy of Sciences. All rights reserved.

chemist and chemical engineer in both academia and industry. This collaboration has been fruitful because both disciplines share a common language and comprehension of physicochemical laws and principles, and a mutual appreciation for the special skills that each brings to bear on the adaptation of chemical discoveries to industrial practice. In contrast, few life scientists are aware of the engineering principles and practical problems associated with the scale-up of biological processes or with the large-scale processing of bioproducts. Conversely, few chemical engineers are sufficiently knowledgeable in the principles of modern molecular and cellular biology, microbiology, genetics, and biochemistry to permit their effective communication with life scientists. This lack of communication inhibits progress on the unique problems that make biological processes particularly difficult to engineer:

- Living organisms can undergo spontaneous mutations, which can have dramatic effects on process operations.
- Processes must be carried out under aseptic conditions to prevent contamination by ubiquitous, undesired organisms.
- Biological processes usually take place in dilute aqueous solution, imposing the need to separate the product from large amounts of water.
- Many of the high-value-in-use products of biotechnology are fragile, hard to purify, and structurally complex.

The training of biochemical engineers should expose them to the power of modern biological science in manipulating and controlling cellular biosynthesis, and to the utility of life sciences techniques and methods in solving critical large-scale bioprocess problems. Prompt and effective exploitation of the "new" biology is critically dependent on the improvement of this life science/biochemical engineering interface; this is one of the most urgent needs confronting biotechnology today.

INTELLECTUAL CHALLENGES AND CRITICAL NEEDS

Because of the rapid progress in the life sciences, commercialization of the fruits of modern biology is not opportunity-limited. The principal limitation is the lack of a technology base that can address the following key challenges:

- Increasing our existing biochemical engineering expertise in techniques of large-scale cell culture not only using simple microorganisms (e.g., bacteria and yeasts) but also utilizing the full range of plant and animal cell systems.
- Expanding our knowledge base so that we will be able to achieve large-scale processing, including the recovery and purification of complex, unstable biological macromolecules (e.g., antibodies, peptide hormones) from product mixtures, as well as the economic and efficient recovery of simple biologically derived organic compounds (e.g., ethanol, amino acids) from highly dilute and impure solutions.

Specific areas of research aimed at meeting these challenges are as follows.

Bioreactors

Mechanically agitated reactors currently used for antibiotic fermentations are often poorly or not at all suited to meet the diverse demands of bioprocessing. New techniques are needed for large-scale culture of plant and animal cells as well as the new microorganisms now being engineered for bioproduct synthesis. Fundamental knowledge of how physical and environmental factors influence intracellular biosynthetic pathways is essential to the development of such techniques. Bioreactor research will require the hybridization of such sciences as molecular and cellular biology, microbiology, and cell physiology with basic engineering skills, chemical kinetics, thermodynamics, fluid dynamics, heat and mass transport, and precise bioprocess

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

control—a rare combination today in either industry or academia.

An important parallel direction of bioreactor research is the development of methods for utilizing free or immobilized enzymes, enzyme combinations, or nongrowing whole cells as catalysts for biosynthesis. The effective utilization of such diverse forms of biocatalysts requires exploration of fluid bed, fixed bed, membrane, cell recycle, tubular, and other reactor types in which to carry out biosynthesis. The challenge is to translate the existing knowledge base for chemical reactors into biosystems, where strict asepsis, complicated biological regulation processes, enzyme and cell fragility, cofactor regeneration of enzyme activity, and maintenance of cell energy provide additional dimensions to the problem. Process fundamentals developed in benchtop systems must be tested in large-scale equipment in order to gain understanding of the scale sensitivity of biosystems.

Separation and Purification

The product stream generated in a bioreactor must go through separation and purification operations in order to isolate the desired product in adequate purity. For fragile, high-unit-value products intended for human or animal health applications, there is a premium on processes that minimize product deterioration and maximize purity. For low-unit-value products where competition lies in nonbiological industrial synthesis the premium lies in recovery processes that are energy-efficient and have high recovery efficiency and low environmental impact. The areas of separation and purification science from which biotechnology can benefit through increased R&D effort fall into three principal categories:

Modification of Conventional, Large-Scale Industrial Separation Methods. Such processes as ion exchange chromatography, pressure-driven membrane separation, electrodialysis, and liquid-liquid extraction are now employed industrially to recover or purify antibotics and simple molecules such as citric acid. They are generally too nonselective or too destructive to be useful for processing fragile protein biologicals. Refinement of these techniques is needed to render them suitable for modern bioproduct recovery.

Adaptation of Biochemical Laboratory Separation Methods to Large-Scale Bioprocessing. Life scientists have developed extremely powerful and sophisticated separation tools such as electrophoretic separation and affinity separation that are potentially applicable to large-scale separations. However, much needs to be learned about the molecular mechanisms and kinetics of these processes before a rational approach to scale-up can be initiated.

Novel Separation and Purification Concepts. There is need for discovery and development of novel separation and purification concepts that embody combinations of physical, chemical, and biological phenomena outside the armamentarium of conventional chemical process technology. Several such processes are being studied in Japan and Europe. Some are based solely on physicochemical principles, for example aqueous two-phase separations employing water-soluble synthetic polymers, or multiple-field fractionation. Other process concepts represent a constructive synthesis of biochemistry and cell biology with industrial chemistry and chemical engineering. Examples include separation based on modification of permeability, separation by selective enzymatic transformation, and separation by genetically manipulated intracellular processes.

Bioprocess Instrumentation and Control

The successful operation of bioreactors and downstream processing equipment requires sophisticated process control, which depends upon accurate measurement of critical process variables and upon the use of advanced estimation algorithms, process models, and

Copyright © National Academy of Sciences. All rights reserved.

control strategies. Current biosensors and control methods often do not provide the desired reliability or the capability for sensitive regulation of process conditions. Furthermore, precise bioprocess control often requires on-line monitoring of the concentration of a complex biological substance for which a specific sensor does not exist. The use of enzymes, monoclonal antibodies, and even whole living cells as components of electrochemical and optical detectors offers promise in solving some of these problems.

Research on bioprocess models is needed to develop optimal control strategies and in order to extract the most useful information from measurements. Formulation of these models will require more advanced control algorithms and a greater knowledge of the effects of engineering parameters on cells and complex molecules.

Trained Personnel

Although the need for trained personnel has already been addressed in this report, one aspect of this problem deserves mention in this section on critical needs. Constructive interaction between life scientists and engineers will be an essential element of innovative bioprocess development. The federal government can make an important contribution at this critical stage in the development of the new bioprocess science and the commercialization of biotechnology by stimulating a research environment in the academic community that will bring together life scientists and chemists with chemical engineers to focus upon bioprocessing needs and problems, and to train a generation of engineers who will be capable of designing, building, and operating the biotechnical industry of the future.

CONCLUSIONS AND RECOMMENDATIONS

If the United States is to capitalize on the revolutionary advances in biology, it must develop both a fundamental knowledge base in chemical and process engineering for biotechnology and a cadre of trained personnel so that U.S. industry can translate fundamental research results into commercial products. The infant U.S. biotechnology industry does not yet have a sufficiently broad product base to support a significant amount of fundamental research in academic institutions. In the short term, the federal government has a critical role to play in attaining the following objectives.

- Establish the fundamental knowledge base to support the design, scale-up, and optimal control of reactors and processes for the large-scale growth of microbial, plant, and animal cells required for full exploitation of the new biology.
- Develop the fundamental knowledge base for large-scale separation and purification processes that will be required to produce the spectrum of potential biochemical products from simple organic molecules to complex proteins.
- Train the next generation of biochemical engineers in a research environment that provides cross-disciplinary exchange of knowledge, industrial collaboration, availability of state-of-the-art facilities and equipment, and support levels that will assure research productivity.

The following recommendations can achieve these objectives if they are implemented at a level that will gain a competitive advantage for the United States. They will assure the development of the needed knowledge base and reduction of the communications gap between life scientists and biochemical engineers. Accordingly, we recommend that the appropriate federal agencies:

- Provide funding at adequate levels for basic research in biochemical and process engineering that will address the research areas of critical importance in this report (see Appendix).
- Establish grants for cooperative, cross-

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

disciplinary research that will stimulate the development of research groups led by joint principal investigators from both the life sciences and biochemical engineering.

- Provide funds for academic units to acquire, operate, and maintain the special types of equipment needed for biochemical engineering research.
- Establish faculty development awards to allow tenured faculty to broaden their expertise in the principles and techniques of biotechnology through special study and leaves, and to provide promising young faculty with the opportunity to immerse themselves in biotechnology research by relieving them from some of their teaching commitments.
- Establish institutional grants to train graduate students, thus providing the personnel that will be needed if U.S. industry is to capitalize on emerging opportunities in biotechnology.

Copyright © National Academy of Sciences. All rights reserved.

Appendix

Research Areas of Critical Importance

A. Bioreactors

1. Quantitative Molecular and Metabolic Reaction Engineering

- *In vivo* control mechanisms and their relationship to metabolic regulation
- Bioenergetics of biochemical cascade reaction networks
- Chemical regulation in biosynthesis
- Kinetics and mechanism of action of genetically engineered microorganisms

2. Fundamental Understanding of the Physical and Environmental Factors in a Bioreactor Needed for Successful Scale-Up

- Fluid dynamic behavior of bioreactors and its influence on biological behavior
- Gas-liquid and liquid-solid mass transfer and heat transfer in bioreactors
- New reactor systems for viscous fermentations
- High-pressure, nonmechanically agitated bioreactors
- Development of perfusion reactor systems that are compatible with large-scale production
- Integration of bioreactor operation with product recovery
- Bioreactors for nonaqueous systems to reduce downstream processing costs
- Containment in bioreactors designed for genetically engineered organisms

3. Expanding the Knowledge Base on Such Biocatalysts as Immobilized Enzymes or Whole Cells

- Mechanistic understanding of enzyme stability and inactivation
- Techniques for immobilizing enzymes and whole cells for large-scale industrial processes
- Bioreactor systems for reactions involving cofactor regeneration
- Transport mechanisms in heterogeneous biocatalysis

4. Engineering Fundamental for the Design of Reactors for Handling Plant and Animal Cells

- Quantitative definition of the stability of, and the effects of mechanical shear on, genetically engineered animal and plant cells
- Physiological characterization of animal and plant cells
- Characterization and fundamental understanding of surface-cell interactions for those animal cells that require attachment to surfaces in order to grow and produce ("anchorage-dependent" cells)
- Development of surfaces for anchorage-dependent cells that can be used in large bioreactors
- Transport phenomena in plant cells due to formation of large cell aggregates

B. Separation/Purification

1. The modification of conventional, large-scale, industrial separation methods to render them more efficient, reliable, and/or economical when employed for biological product recovery. Examples in this area include:

- Column chromatography: development of new chromatographic materials that do not alter the structural characteristics of biological macromolecules
- Pressure-driven membrane separation processes: membranes that are less subject to fouling by biological process streams
- Liquid-liquid extractors: development of liquid two-phase systems that do

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

not alter the structural integrity of biological macromolecules
- Electrodialysis: development of membranes with selectivities useful for product streams from the "new" biology

2. Adaptation of Biochemical Laboratory Separation Methods to Large-Scale Bioprocessing
- Electrophoresis: Protein behavior in combined electrical and low gravitational fields as well as in centrifugal fields and in flow fields in gradient pore-size gels
- Affinity chromotography: mechanisms and kinetics of antibody-antigen interactions
- Affinity chromotography: factors that influence the association/dissociation equilibria in affinity separations

3. Novel Separation and Purification Process Concepts
- Fundamental principles in aqueous two-phase separations
- "Multiple Fields" such as electrical, magnetic and gravitational fractionation processes
- Separation principles and technologies based on selective enzymatic transformations
- Biological and chemical alteration of cell-wall permeability to facilitate product separation
- Integration of modern biology with downstream recovery science to alleviate potential purification problems by involving biochemical engineers in the design of recombinant cells to perform the biosynthesis

C. Bioprocess Instrumentation and Control

1. Measuring Instruments for Important Biological Variables:
- Sensors for viable cell density, biological product concentrations, cellular metabolic state, and mixed culture compositions
- More reliable and sensitive instruments for measuring concentrations of dissolved oxygen and carbon dioxide, ionic strength, and other abiotic parameters
- Biosensors based on enzymes, antibodies, and cells with adequate stability and durability

2. Mathematical Models to Aid in Process Design, Control, and Optimization:
- Process descriptions based on key features of biological mechanisms rather than on empirical correlations
- Models that accurately represent the effects of simultaneous changes in several process variables

3. Advanced Control Strategies for Regulation of Complex Biological Processes:
- Data analysis and interpretation algorithms based on biological structure and statistical filtering principles
- Control for new multiphase, multifunction bioprocesses (one example would be a bioprocess combining the reaction step with the product separation step)
- Methods for identifying and counteracting undesired cellular control mechanisms
- Controls to maximize the useful lifetime of immobilized biocatalysts and separation media

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Report of the Research Briefing Panel on High-Performance Polymer Composites

Copyright © National Academy of Sciences. All rights reserved.

Research Briefing Panel on High-Performance Polymer Composites

David McCall (*Co-Chairman*), AT&T Bell Laboratories, Murray Hill, N.J.
Rudolph Pariser (*Co-Chairman*), E. I. du Pont de Nemours & Co., Inc., Wilmington, Del.
Paul J. Flory, Stanford University, Stanford, Calif.
Frank Gerstle, Sandia Laboratories, Albuquerque, N. Mex.
Allan S. Hay, General Electric, Schenectady, N.Y.
Ted Helminiak, Wright-Patterson AFB, Ohio
Frank Kelley, University of Akron, Akron, Ohio
William J. MacKnight, University of Massachusetts, Amherst, Mass.
Arthur B. Metzner, University of Delaware, Newark, Del.
Seymour Newman, Ford Motor Company, Detroit, Mich.
Manuel Panar, E. I. du Pont de Nemours & Co., Inc., Wilmington, Del.
Darrell H. Reneker, National Bureau of Standards, Gaithersburg, Md.
L. E. Scriven, University of Minnesota, Minneapolis, Minn.
Leslie Smith, National Bureau of Standards, Gaithersburg, Md.
George M. Whitesides, Harvard University, Cambridge, Mass.
Dick J. Wilkins, General Dynamics Corporation, Fort Worth, Tex.

Staff

William Spindel, *Study Director*
Robert M. Simon, *Staff Officer*
Robert M. Joyce, *Rapporteur*
Robin Illsley, *Secretary*

Copyright © National Academy of Sciences. All rights reserved.

Report of the Research Briefing Panel on High-Performance Polymer Composites

EXECUTIVE SUMMARY

High-performance polymer composites are formed from exceptionally strong, highly oriented, long fibers, such as graphite, aramid, or glass, which are placed in a desired orientation and bound together by a polymer matrix. The strengths and moduli thus achieved are on a par with those of the strongest available structural metals, but are substantially higher on an equal-weight basis.

The utility of advanced composites is based on their anisotropic physical properties and their unique strength and modulus per unit weight. These properties have led to significant adoptions as structural members and skins of military aircraft. As experience has been gained in the design and fabrication of these materials, they are finding increasing use in space vehicles, civilian aircraft, recreational equipment, and moving parts in textile machinery. Large-scale applications in automobiles, heavy equipment, robotics, and building and bridge construction are foreseen. We are now on the threshold of realizing the vast potential of advanced composites, which promise to play a major role in the U.S. economy and defense.

The field of advanced composites is an international arena. The present U.S. position is strong in chemistry, materials engineering and application, but Japan dominates in carbon fiber technology. In Japan, a broad MITI program covering advanced composites is in place. In West Germany, a new Max Planck Institute for polymer research will impact on composite technology.

Research needs include the following areas:

- New fiber compositions, polymer matrices, and fabrication methods are needed to broaden applications and to reduce costs.
- Understanding of composite structure/property relationships is lacking. The fiber/matrix interphase region has not been well characterized, although it is critical to composite performance.
- Knowledge of failure mechanism is primitive. Flaw identification, growth, and elimination need study. Response to differing load conditions and environmental factors has not been described in fundamental terms. Life prediction, reliability, and testing are critical matters that await characterization and understanding at the molecular level.
- A science of composite design and pro-

Copyright © National Academy of Sciences. All rights reserved.

cessing needs to be developed and must involve extensive computer-based modeling for design, engineering, and manufacture. Joining and repair processes must also be developed on a fundamental basis. The toxicity, envionmental effects, and eventual disposal of composites must be examined for their consequences on fabriciation and use of these materials.

• Reinforcement of a polymer matrix at the molecular level, with stiff, rodlike polymer molecules, has been demonstrated and should be studied to learn how to develop optimum properties in such composites.

The rapid technological advances in high performance composites over the past decade have outpaced the development of the underlying basic science. Research in the area calls for communication and collaboration among several disciplines, including chemistry, physics, chemical engineering, mechanical engineering, and materials science. These groups tend to be insular in present U.S. university structures. The future will belong to those who can bring together the necessary disciplines and construct a coherent intellectual structure that will form the basis for advancing to new levels of understanding.

Research on advanced composites in U.S. universities is embryonic and frequently includes only one of the several science and engineering disciplines needed to advance the subject. A realignment of the traditional academic structure appears necessary if the needed interdisciplinary research is to be effective. Fewer than 30 U.S. universities conduct research on composite materials, involving about 40 equivalent full-time faculty. Multidisciplinary organizations devoted to composites exist at only two of these universities. Moreover, a substantial fraction of this university research is devoted to product design and to lower-performance short-fiber composites.

Estimates of overseas support for composites research indicate that several countries have efforts that are comparable to that of the United States. However, their programs are growing more rapidly than ours and are concentrated on the interdisciplinary approach that is essential to progress.

It is recommended that a modest number—perhaps three, as it becomes feasible to staff them—of interdisciplinary centers devoted to basic research on advanced composites be established in universities. The proposed Engineering Research Centers are a good model, and research on composites should be encouraged in them. In addition, more research specifically concerned with advanced composites should be encouraged in existing academic interdisciplinary centers devoted to materials science. The objectives of these centers would be to develop the scientific underpinning that is essential to advance the national position in advanced composites and to develop a pool of trained scientists and engineers for future personnel needs in research, development, and teaching in this high-technology area. Industrial involvement in the form of intellectual presence and research cooperation in the campus centers is regarded as essential.

INTRODUCTION

Structural composites are not new. Nature has been perfecting these materials for millions of years, resulting in such composite structures as wood, bone, or clam shells. These structures comprise combinations of components specifically chosen in terms of amount, size, shape, orientation, and interface to achieve a set of overall properties that is optimized for its intended purpose.

This is also the central purpose of man-made high-performance composites, a relatively new area of technology that holds promise for revolutionary advances in structural materials. These composite materials are filling crucial national needs. They are already essential to the defense establishment and the

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

space program, and are beginning to have significant impact in important areas of the national economy, particularly aircraft, recreational products, textile machinery, and ground transportation.

Precursors to advanced composites are found in such diverse developments as the automobile tire and short-fiber reinforced plastics. In recent years, several classes of high-performance structural composites have been developed, including multicomponent ceramics, short-fiber reinforced metals, and polymer composites based on strong, highly oriented, continuous fibers developed through advances in inorganic and polymer chemistry. High-performance polymer composites are made by arranging arrays of strong, stiff fibers in a desired orientation, usually in close-packed alignment, and binding them together with polymer matrices. This new class of polymer composites provides combinations of strength and modulus substantially superior to those of structural metals and alloys on an equal weight basis. This report is concerned only with these high-performance polymer composites.

The principal impetus for the development of high-performance polymer composites came from military aircraft and space vehicles, which needed structural components of minimum weight and with high strength in prescribed directions. Further impetus was provided by the adoption of these composites in civilian aircraft to achieve fuel savings, in recreational equipment, and in moving parts of textile machinery to achieve higher operating speeds. We are now only on the threshold of realizing their vast potential. Significant improvements in cost and performance are expected and should greatly expand their uses.

The United States pioneered the use of advanced composites and is currently in a strong position in this technology. However, significant programs on composites research are being implemented in other countries. Specifically, a broadly based academic-industrial program, initiated by MITI, is in place in Japan. Germany's composites research will likely be strengthened soon by a new Max Planck Institute for polymer science and engineering in Mainz.

The key issue addressed in this report is the need for increased research in basic science and engineering in high-performance composites to provide the technical foundation for meeting important national needs. The rapid technological advances in high-performance composites over the past decade have greatly outpaced the development of the underlying basic science. Further improvement in such areas as composite manufacturing and processing will undoubtedly emerge from continuing applied research and development. However, basic research in this area must be strengthened and broadened to acquire needed knowledge and to assure a supply of professional scientists and engineers with the background necessary to advance the technology. An investment in basic research would be a high-leverage investment.

This report is organized into the following sections: current and projected APPLICATIONS of high-performance composites; SCIENCE AND TECHNOLOGY—the present and prospective state of the art, international competition, and goals and directions for research; EDUCATIONAL NEEDS; and CONCLUSIONS AND RECOMMENDATIONS.

APPLICATIONS

The evolution of high-performance composites is founded in the opportunity to build light-weight, high-strength, high-modulus structural parts that can be tailored to meet specific loading conditions. Because these composites can be made with superior strength and modulus in one or two directions, structural parts can be designed specifically to accommodate the anticipated direction of loading, in contrast to isotropic

Copyright © National Academy of Sciences. All rights reserved.

metallic parts. Advanced composites are currently high in cost, reflecting costs of fibers and manufacturing processes; both cost elements are decreasing and should continue to do so.

High-performance composites are widely used in space vehicles because they lower launch weight and have good dimensional stability over a wide range of temperatures owing to their low coefficients of thermal expansion. These composites are used as structural components in many military and civilian aircraft and will comprise a larger fraction of the structural members in the coming generation of aircraft. Range, payload, fuel consumption, tooling costs for parts, and radar profile all enter into design considerations, and the payoff can be enormous. Japan has identified the aircraft industry as one of five in which they expect to be a leader in the 1990s, and their commitment to composites research is part of their strategy.

Sporting goods, including spars for boats, racing car bodies, bicycle frames, tennis racquet frames, fishing rods, and golf club shafts, bring high-performance composites to the public eye. Composites provide superior performance, and users are willing to pay for it.

Use of advanced composites in automobiles and other surface vehicles has been limited, and a number of problems must be solved before this industry can realize the benefits of these materials. An acceptable balance between processing speed and product quality has not yet been achieved. Useful technology for joining and repairing is not in hand, and long-term dimensional stability must be improved. However, there are important benefits to be gained by using composites in such vehicles. Tooling costs are much lower than for steel, providing greater manufacturing flexibility, more rapid design changeover, and reduced capital investment. It has been estimated that the 1,500 body parts in one automobile could be reduced to a substantially lower number by substituting high-performance composites for metals. Composites are less subject to corrosion than metals, and vehicle weight is reduced, with consequent fuel savings. At present, advanced composites are finding applications in leaf springs and drive shafts in production vehicles, and experimental composite engines have been constructed and tested. Automotive applications in the United States could expand dramatically in response to lower composite costs and the development of lower-cost manufacturing processes for using them.

Military uses in addition to aircraft are many and growing. As in sports equipment, performance rather than cost is the driving force. Examples include helmets, body armor, parts for personnel vehicles and for tanks, light-weight segments for movable bridges, carriages and braces for artillery, and portable rocket launchers.

High-performance composites have great promise in industrial machinery as reciprocating, oscillating, or rotating parts where the weight, strength, fatigue life, noise, and natural vibrational frequency of metals limit operating speeds and thus productivity. These composites are being used for moving parts in textile machinery, where higher operating speeds and productivity are paramount. A carbon fiber/epoxy centrifuge rotor can be spun at twice the speed of one made of high-tensile steel because of its high specific strength, providing a separation efficiency four times that of the steel rotor. Significant productivity increases can be foreseen by taking advantage of the high specific strength and modulus, lower moment of inertia, and the self-damping characteristics of composite parts.

High-performance composite parts have potential uses in robots and heavy construction machinery, where the low moment of inertia of a cantilevered member translates into lower energy costs and more accurate control of its positioning. Another potentially large area of use is load-bearing members of buildings and bridges, where the strength/weight ratio, self-damping, and corrosion resistance offer substantial benefits.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

In summary, advanced composites are being used in a number of industries that are vital to the national economy. Their uses in these industries will unquestionably expand with anticipated improvements in cost, manufacturing speed and consistency, and particularly in basic understanding of the critical properties that must be measured to ensure optimum design and prediction of the useful life of structural parts.

SCIENCE AND TECHNOLOGY

FIBERS

High-strength fibers are the primary load-bearing components of high-performance composites; they comprise about 60 percent of the volume of the composite and largely control both properties and cost. Carbon, aramid, and glass are the most widely used fibers. They have tensile strength/density and modulus/density ratios several times that of steel and represent a new technology that is just beginning to demonstrate its potential.

Glass fibers were the basis for some of the earliest reinforced polymers. The development of composites of continuous glass fibers, although largely empirical, led to the first controlled processing conditions and microstructure that took advantage of the fiber strength.

Carbon fibers are stronger, stiffer, and more costly than glass. They are produced by pyrolyzing oriented polyacrylonitrile to leave carbon, which is graphitized. The process may include hot stretching. Lower-performance carbon fibers are made from coal tar or petroleum pitch. Empirical adjustment of processing parameters has resulted in commercial carbon fibers in which either tensile strength or modulus is optimized. The strain of carbon fibers at break is about 1 percent.

Aramid fibers are prepared by spinning a polymer that has a highly rigid linear molecular structure from an oriented anisotropic phase, in which the molecules are spontaneously oriented over microscopic dimensions. During spinning the oriented regions are aligned with the fiber axis to give high strength and rigidity. Because of the low density of the polymer (1.44), the specific tensile strength of the fiber is more than 5 times that of steel. The elongation at break is greater than that of nonmetallic fibers of similar modulus, and this property can be an advantage in composite design because toughness is related to energy absorption before break. Aramid fibers have excellent thermal properties (although inferior to those of inorganics), good chemical stability, and low flammability. Like carbon, they can be made in a range of properties that optimize either strength or modulus.

The spinning of an anisotropic phase has been extended to both solutions and melts of newer polymers. For example, experimental samples of polymers such as polybenzothiazole appear to have better properties for composite use than the commercial aramids. The spinning of anisotropic melts offers the potential of lower costs than spinning from solution, but the polymers that can be spun from melts tend to give fibers with lower use temperatures. Another example is polyethylene, which usually has low modulus and strength. It has been fabricated into ultra-high-strength fibers by gel spinning. Processes for making fibers from poly(ethylene terephthalate) have recently been patented by a Japanese group.

Boron fibers are prepared by decomposition of a volatile boron compound on a resistively heated fine tungsten filament, which is converted into tungsten boride and serves as the fiber core. Fibers of silicon carbide can be made by a similar process or by pyrolysis of a polycarbosilane fiber. Fibers of silicon nitride and alumina have also been developed.

Since the various fibers have different mechanical properties as well as costs, it is not uncommon to create hybrid composite structures comprising more than one kind of fiber to balance performance/cost requirements.

Tensile failure in carbon, glass, and other inorganic fibers is typically initiated by a flaw

Copyright © National Academy of Sciences. All rights reserved.

TABLE 1 Properties of Fibers

		Steel	Aluminum	E-Glass	S-Glass	High-Strength Graphite	High-Modulus Graphite	Aramid	Boron
Tensile strength (10^3 psi)	Achieved 1984	500	100	350	650	720	350	525	500
	Theoretical	3,000	1,000	1,000	1,280	6,000	6,000	4,000	5,800
Tensile modulus (10^6 psi)	Achieved 1984	30	10	10	12.8	36	75	18	58
	Theoretical	30	10	—	—	100+	100+	29	—
Density (g/cm^3)		7.8	2.7	2.54	2.5	1.75	1.8	1.44	2.7
Thermal expansion (10^{-6}/°c)		11	24	5	3	−1	−1	−2	5
Specific strength (10^6 in.)		1.8	1.0	3.9	7.3	11.5	5.4	10.5	5.2
Specific modulus (10^8 in.)		1.1	1.0	1.1	1.4	5.8	11.7	3.5	6.0

or surface growth step that causes local stress concentration and produces a simple fracture perpendicular to the fiber axis. The weak lateral bonding and oriented polycrystalline nature of organic fibers such as aramid produce a fibrillar failure. Considerable splitting along the fiber axis occurs and results in a mixed shear and tensile failure.

In compression, carbon fibers usually fail by combined interface separation and microbuckling. Aramid develops shear or kink bands that form at relatively low stresses but do not produce catastrophic failure. Glass and boron fibers fail in compression when an axially oriented flaw is activated by expansion in the noncompressed direction. Compressive strengths of boron and glass composites, as well as those made from alumina and silicon carbide, exceed their tensile values.

The origin, growth, and interaction of fiber flaws with the composite matrix are poorly understood. Failure mode is not predictable from tensile strength and modulus, and the question of what measurable properties are critical to behavior in use has not been resolved. It is recognized that the time-dependent behavior of fibers under stress is important. Most test numbers are derived from short-term static tests, which do not give a reliable indication of long-term life in use.

The strength and modulus of current carbon, aramid, and glass fibers are signficantly below the calculated theoretical values. However, substantial improvements in the properties of all three have been made over the last 10 years; while it is unlikely that the theoretical properties will be attained, there is good reason to expect that improvements will continue to be made. Moreover, the possibilities in organic polymers for new high-strength, high-modulus fibers have only begun to be explored, and it is expected that organic fibers with strengths 2-3 times those currently available will be made.

Matrices

The matrix component of a composite serves primarily to bind the load-bearing fibers. The matrix is the minor component of the composite, and, while it does not usually support stress, flaws or damage in it may lead to failure of the composite. The properties of the matrix usually determine the limits of temperature and chemical environment that the composite can withstand.

Current polymer matrix materials are principally thermosetting epoxy resins, which are chemically cured to produce highly cross-linked structures. The curing process requires elevated temperatures and substantial times. Cross-linked epoxies have very low fracture toughness, and their composites are intolerant of damage because cracks will propagate readily through the matrix. Tougher materials have been developed by incorporating a rubbery dispersed phase in the epoxy, but

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

these polymers have lower continuous use temperatures and lower modulus. Epoxy composites can be used at temperatures up to about 175°C. However, under hot/wet conditions they lose properties at lower temperatures. New high-temperature/high-modulus resins, such as polyimides, bismaleimides, and acetylene-terminated oligomers based on high-temperature resins, promise improved matrix properties with lower moisture sensitivity and faster curing. These matrices offer the potential of use temperatures in the range 200-370°C. However, their toughness is still low, and they must also be cured for significant times at elevated temperatures.

Tough, high-temperature-resistant and solvent-resistant thermoplastics with low moisture sensitivity are being explored as matrix resins. Examples are polyamide-imides, polyphenylene sulfide, polyetherimides, and polyether-etherketone (PEEK). They can be processed in the melt or in solvents that must subsequently be evaporated. They are strong candidates for replacing thermosets because of their shorter fabrication cycle, which does not require a chemical cure. Property improvements should result from their greater inherent toughness and the fact that some of them are less water-sensitive than thermosets. Thermoplastics creep under stress at elevated temperatures, and current materials have use temperatures in the range 150-230°C.

The ideal resin for a matrix material would be a low-viscosity material that could easily be combined with the reinforcing fiber and could then be converted rapidly and reproducibly at low temperatures into the final composite structure. Important properties of the matrix include high and/or low temperature capability, resistance to water and solvents, toughness, and a coefficient of thermal expansion that matches that of the fiber. Although it is unlikely that all of these properties will be found in one matrix material, the potential of new polymer-forming monomers and reactions, copolymers, block copolymers, and blends should lead to a variety of high-performance matrices.

Composites

One method of manufacturing high-performance composites involves the resin-impregnation of bundled continuous filaments of a reinforcing fiber to form a continuous tape or fabric. This prepreg is subsequently plied in layers that may have the same or different alignments. The plied structure is then subjected to pressure and heat, which yields a consolidated composite structure for subsequent assembly or manufacturing steps. The combination of structural design efficiency and the flexibility of the manufacturing process through the layering of directional plies provides the desired structural properties.

Another manufacturing process involves winding a coated fiber on to a mandrel. A computer-controlled payout system capable of multiaxis precision winding can generate virtually any geometry on an irregularly shaped mandrel. The fiber-wound structure is impregnated with polymer, which is then cured, and the composite structure removed from the mandrel.

Although some aspects of composite technology are quite advanced, in others our scientific understanding and processing methods are clearly inferior to, for example, the knowledge of structure-property relationships and fabrication techniques for steel or aluminum alloys. Since structural metals are the principal competition for composites, increased acceptance of high-performance composites depends on their becoming as reliable, simple, and economical to produce as highly developed alloys.

High-performance composites have presented engineers and materials scientists with a variety of new challenges. Analytical and numerical techniques had to be developed to treat anisotropic mechanics problems. In particular, linear elastic mechanics needed for orthotropic plates and shells, and

Copyright © National Academy of Sciences. All rights reserved.

the micromechanics for calculating homogeneous, thermomechanical properties are now in hand. Techniques also exist for predicting thermoelastic properties and for detecting mechanical or compositional flaws for the more common composite components. Processing and fabrication of thin laminates into consistently high-quality materials have become routine.

However, composite manufacturers have not yet developed techniques for consistent production of thick laminates for highly stressed structural elements; fibers often move during the fabrication process, with deterioration in composite properties. Producing a void-free, high-fiber-volume laminate still involves much empiricism, since the process of creating and curing a laminate is difficult to model. Aspects of high-performance composites that are less developed, but also needed for design, are data on dynamic tensile and fatigue strength. Progress has been made in predicting the effect of flaws on failure modes dominated by the resin, but not those dominated by the fiber.

COMPETITIVE WORLD POSITION

The United States has a strong position in aramid fibers and a reasonable level of activity in developing other organic fibers with comparable or better properties. However, aramid fibers may be soon commercialized in both Europe and Japan, and the activity on other organic fibers appears to be at least as strong in Japan. MITI has designated the development of a "third-generation" fiber as a government-subsidized project, beginning in 1983 and targeted for practical applications in 5 years. Because of the overwhelming importance of the load-bearing fibers in composites, the field is sensitive to breakthroughs in stronger fibers. Current patent activity suggests that such a breakthrough is as likely to come from Japan as from this country. For example, processes to make high-strength fibers from poly(ethylene terephthalate) have recently been patented by a Japanese group.

Much of the technology used for manufacture of carbon fibers in the United States is licensed from Japanese companies. Although there is research activity in this country, some companies have chosen joint ventures with Japanese firms to take advantage of Japan's developed technology. The high level of Japanese carbon fiber technology also suggests that they may produce many of the expected future advances in these materials.

The United States currently has a strong position in composite manufacturing and processing technology and leads the world in developing major applications in such industries as aircraft, sporting goods, and automotive. However, there is growing oversees activity in composites technology. Foreign manufacturers are gaining valuable experience by subcontracting for the manufacture of large structural parts for new-generation U.S. commercial aircraft, using technology supplied by the United States. European countries are as active as the United States in putting high-performance composite parts into their military aircraft. Japanese and West European technologies are beginning to rival ours.

RESEARCH DIRECTIONS

The rapid technological development of advanced composites has outpaced the basic understanding of the structure and performance of these materials. A strengthening of basic research in this area is needed to provide a firm scientific foundation for the anticipated future advances in composite technology. General directions for such research are summarized below.

We need better understanding of molecular structure/property relationships of fibers and of how such properties translate into their behavior in a composite. Composites are subjected to long-term cyclical loading, and we need to understand the causes of fiber

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

fatigue under dynamic conditions. Although there is some knowledge of how different fibers fail under tensile and compressive loads, there is essentially no knowledge of how such failure may be dictated by the fiber microstructure. We also need to investigate the effects on composite properties of fiber molecular structure and macrostructure: molecular composition, molecular orientation, surface structure, size, and geometrical placement in the composite, and how composite properties are affected by processing.

We need to define the physical and chemical parameters of matrix resins that can be translated into prediction of performance in a composite. The curing of thermosetting matrices should be studied to determine the effects of curing conditions on the cross-linked network structure, physical properties, and fiber interactions with the cured resin. Structural modifications of epoxies and other thermosets that will increase their toughness should be explored. Thermoplastic polymers have potential advantages over thermosets in offering greater toughness and requiring shorter composite fabrication times. The interaction of thermoplastic polymers with fibers at the interface will be substantially different from that of thermosets and need to be understood at the molecular and morphological levels. Research may point the way to reducing the tendency of thermoplastics to creep at elevated temperatures under load.

There is much to be learned about the fundamentals of the fiber-polymer interface: How does adhesion affect the load bearing, and how can optimum adhesion be achieved; what are the molecular structure and mobility of the polymer at the interface; what can be learned about nonequilibrium polymer conformations at the interface; how does the interface affect the kinetics of growing polymer chains; what properties can be achieved with various polymers and fibers when the interface is prepared under ideal laboratory conditions?

We lack the knowledge base for combining fibers and matrices to take maximum advantage of their properties. Nondestructive test methods must be developed to determine whether the manufacturing process has attained the desired microscopic and macroscopic structures, with the fibers in their desired positions. We need more precise techniques for predicting composite properties from properties of components, for detecting and predicting effects of flaws, and for predicting remaining service life of a part in use. Because of our limited basic knowledge, composite structures are now designed with considerably greater margins of safety than those used for metal structures, thereby losing some of the inherent superiority of composites. We need better techniques for joining or fastening composite parts to metals with minumum adverse effects on composite strength.

Many potential applications for advanced composites will depend on achieving substantial improvements over current manufacturing speeds and product quality. We must develop automated robotic manufacturing processes to reduce product variability. Composite samples prepared under carefully controlled laboratory conditions should provide benchmark properties against which to assess manufactured parts.

The toxicity of composite components, their long-term environmental effects, and their recyclability need to be studied, and should be considered at the development stage of a composite.

Finally, we should draw upon knowledge of the structure of natural composites such as bone, wood, and clam shells for clues to how man-made composites can be made to approach more nearly their theoretical physical properties.

EDUCATIONAL NEEDS

The rapid technological advances in high-performance composites over the past decade

Copyright © National Academy of Sciences. All rights reserved.

have outpaced the development of the underlying basic science. Research in the area calls for communication and collaboration among several disciplines, including chemistry, physics, chemical engineering, mechanical engineering, and material science. These groups tend to be insular in present U.S. university structures. The future will belong to those who can bring together the necessary disciplines and construct a coherent intellectual structure that will form the basis for understanding present achievements and advancing to new levels.

Research on advanced composites in U.S. universities is embryonic and frequently includes only one of the several science and engineering disciplines needed to advance the subject. A realignment of the traditional academic structure appears necessary if the needed interdisciplinary research is to be effective. Fewer than 30 U.S. universities conduct research on composite materials, involving about 40 full-time equivalent faculty. Multidisciplinary organizations devoted to composites exist at only two of these universities. Moreover, a substantial fraction of this university research is devoted to product design and to lower-performance short-fiber composites. Consequently, too few scientists and engineers are being trained with the broad background needed to advance composite technology, and the gap between supply and demand for trained people is increasing with the rapid growth of the composites industry.

Estimates of overseas support for composites research indicate that several countries have efforts that are comparable to that of the United States. However, their programs are growing more rapidly than ours and appear to be concentrated on the interdisciplinary approach that is essential to progress.

CONCLUSIONS AND RECOMMENDATIONS

Our conclusions are as follows:

- High-performance composites will find increasing numbers of applications in both industry and defense. They are essential to many high-technology industries and they will meet important national needs.
- The present U.S. position in high-performance composites is strong in the areas of chemistry, materials engineering, and applications. There is a strong challenge to the United States in this area from both Japan and Europe.
- Basic research on high-performance composites in the United States must be strengthened to develop the understanding needed to advance the field and to maintain the U.S. position against foreign competition.

It is recommended that a modest number—perhaps three, as it becomes feasible to staff them—of interdisciplinary centers devoted to basic research on advanced composites be established in universities. The proposed Engineering Research Centers are a good model, and research on composites should be encouraged in them. In addition, more research specifically concerned with advanced composites should be encouraged in existing academic interdisciplinary centers devoted to materials science.

The objectives of these centers would be to develop the scientific underpinning that is essential to advance the national position in advanced composites and to develop a pool of trained scientists and engineers for future personnel needs in research, development, and teaching in this high-technology area. Industrial involvement in the form of intellectual presence and research cooperation in the campus centers is regarded as essential.

Copyright © National Academy of Sciences. All rights reserved.

Report of the Research Briefing Panel on the Biology of Oncogenes

Copyright © National Academy of Sciences. All rights reserved.

Research Briefing Panel on the Biology of Oncogenes

Robert A. Weinberg (*Chairman*), Associate Professor of Biology, Massachusetts Institute of Technology, Whitehead Institute, Cambridge, Mass.
Raymond Erikson, American Cancer Society Professor of Cellular and Developmental Biology, Biological Laboratories, Harvard University, Cambridge, Mass.
Fred Rapp, Professor and Chairman, Department of Microbiology, The Milton S. Hershey Medical Center, Pennsylvania State University, Hershey, Pa.
Michael Sporn, Chief, Laboratory of Chemoprevention, National Cancer Institute, National Institutes of Health, Bethesda, Md.
George Todaro, Scientific Director, Oncogene, Seattle, Wash.
George Vande Woude, Director, LBI Basic Research Program, NCI Frederick Cancer Research Facility, Frederick, Md.
Michael Wigler, Senior Staff Scientist, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Rapporteur

Luis F. Parada, Massachusetts Institute of Technology, Whitehead Institute, Cambridge, Mass.

Staff

Barbara Filner, *Director,* Division of Health Sciences Policy, Institute of Medicine
Barbara Mandula, *Consultant*
Naomi Hudson, *Administrative Secretary*

Copyright © National Academy of Sciences. All rights reserved.

Report of the Research Briefing Panel on Biology of Oncogenes

ONCOGENES AND THE MOLECULAR BASIS OF CANCER

Introduction

Researchers have sought to understand the causes of cancer for many decades. The advances prior to the early 1970s were largely descriptive. However, over the last 10 years, new ideas have been formulated and better technologies developed to address the problem. Possibly the most important result of these labors has been the recent exciting discovery of cellular oncogenes, or cancer genes. Cancer no longer appears as a vaguely defined collection of phenomena with no central mechanistic basis. Researchers are now beginning to define in precise terms—down to the last molecule—the kinds of changes that occur within a normal cell to propel it forward to becoming a tumor cell.

Various agents from outside the organism, such as radiation, chemical substances, and certain types of viruses, have been identified as involved in cancer induction. The accumulated evidence now suggests that viruses and radiation are involved in causing only a minority of human cancers. The suggestion that chemical substances (including agents in tobacco smoke and in the diet) induce the bulk of human cancers did not itself shed light on the molecular processes involved in the transformation of normal cells into tumor cells. A central question remained: How does a cancer-causing chemical alter a cell so that this cell and its descendants exhibit cancer traits? It appeared that the genetic blueprint of these cells, carried in their DNA, must be permanently changed (mutated). This implicated DNA molecules as the central target of carcinogenesis.

The chasm between the formulation of this explanation and a demonstration of its correctness could be bridged only by recent advances in cellular and molecular biology. These advances, both conceptual and technical, led to the discovery of cellular oncogenes and some of the molecular mechanisms of carcinogenesis. Modern DNA technology, developed in studies seemingly unrelated to cancer, has been mustered to isolate specific DNA sequences, analyze them in great detail, and begin to relate their properties to the process of carcinogenesis.

Copyright © National Academy of Sciences. All rights reserved.

The Search for Oncogenes

Retrovirus-Associated Oncogenes

Our initial knowledge of cellular oncogenes stemmed from tumor virology. RNA tumor viruses ("retroviruses") were known to cause cancer in lower animals. These viruses harbor cancer-causing genes, called viral oncogenes, that are not genuine viral genes, but rather genes from animal cells that were captured by the viruses in the course of infecting cells. These acquired, previously normal, cellular genes fell under the control of the viral genetic machinery. The result was that a normal cellular gene, a "proto-oncogene," was converted into a potent oncogene. There are more than 20 different oncogenes now known to be associated with a variety of retroviruses, each pointing to the existence of an antecedent proto-oncogene residing among the genes of normal cells.

The normal human genome has precise counterparts (homologues) to the proto-oncogenes originally found in animal genomes. Proto-oncogenes also have been detected in such other organisms as yeast, fruit flies, and mice. The conservation of these proto-oncogenes during evolution suggests that they play a crucial role in cellular and organismic survival.

Discovery of Human Oncogenes

Gene transfer or "transfection" techniques make it possible to extract genes from one cell and introduce them into another. Transfection experiments showed that DNA segments from human cancer cells contain genetic information able to transform normal recipient cells into tumor cells. These DNA molecules contained specific transforming segments termed, once again, oncogenes. Such oncogenes were found in a wide variety of human tumors having no known viral etiology. All the isolated oncogenes were very similar to genes found in normal cells. This confirmed the suspicion that *oncogenes of human tumor cells arise by alteration of genes present in normal human cells.*

Subsequent results were surprising in two ways. First, the human oncogenes were closely related in structure to the oncogenes that have been acquired and activated by cancer-causing animal retroviruses. Thus, mechanisms of cancer causation in humans and animals apparently have common molecular features. By extension, all types of cellular proto-oncogenes that had been activated by animal retroviruses could potentially also be activated in human cancers.

The second surprise occurred when researchers first compared the molecular sequence of a human cancer gene, isolated by recombinant DNA technology, with that of the corresponding proto-oncogene that resides in a normal cell. They found that this oncogene, present in a human urinary bladder carcinoma, had become activated by the mutation of just one nucleotide among the 5,000 nucleotides comprising the normal proto-oncogene. This change resulted in one different amino acid in the protein encoded by that gene. Thus, a very limited change in the structure of a normal protein resulted in a new mandate for the cell that harbored it. That first evidence—now much expanded—showed that important steps in carcinogenesis can be traced to specific, well-defined genetic lesions. Presumably, the mutation responsible for the activation of this particular proto-oncogene occurred in bladder tissue during the course of tumor formation.

About 15 to 20 percent of all human tumors, from a wide variety of body sites, have been shown to carry oncogenes in their DNA. Most of these oncogenes are members of the *ras* gene family. (Oncogenes are given three-letter names, such as *ras, src, myc, mos; ras* stands for rat sarcoma, the tumor in which it was first found.) Some of the *ras* genes have been activated in experimental animals with x-radiation and with chemicals known to be mutagenic and carcinogenic. Transfection assays continue to unveil new human onco-

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

genes, at least six of which are now being characterized in laboratories around the world.

Oncogenes and Chromosomal Abnormalities in Malignancies

Many types of human cancers carry specific chromosomal rearrangements. The role of these rearrangements in cancer development was not understood until molecular analysis of human proto-oncogenes became possible.

The best examples of specific chromosomal rearrangements are the "translocations" found in human Burkitt's lymphoma and in mouse plasmacytomas. In both of these diseases, a chromosomal rearrangement relocates the normal *myc* proto-oncogene from its usual chromosomal locus into the vicinity of antibody-producing genes. As a result of this chromosomal translocation, the previously normal *myc* gene is no longer regulated by its usual control mechanisms. With loss of regulation, the *myc* gene is converted into an active oncogene that contributes to the creation of a tumor cell.

In addition to translocations, cancer cells often carry genes in amplified copy number (many copies of the same gene). DNA analysis of some of these amplified genes shows that they are oncogenes. The N-*myc* oncogene is amplified in advanced stages of human neuroblastoma; the *ras* genes have been detected in amplified form in several human cancers; and the *myc* oncogene is amplified in promyelocytic leukemia, colon carcinomas, and small cell lung carcinomas. The amplified copies of these genes cause overproduction of gene-associated products.

Biochemical Characterization of Oncogene Proteins

Oncogenes exert their effect on cells by means of the proteins whose production is specified by the nucleotide sequences of the genes. The amino acid sequences of almost all the known oncogene proteins have been determined. In many cases, the oncogene proteins closely resemble normal cellular proteins. This close relationship suggests that the oncogene proteins function similarly to normal cellular proteins, but in a deregulated or inappropriate way.

The best characterized of the oncogene proteins is that specified by the *src* oncogene. This protein, associated with the inner surface of the cell membrane, acts as a protein kinase; it attaches phosphate groups covalently to other proteins or to itself. These phosphate groups are linked to the tyrosine amino acids of the target proteins. This specificity distinguishes oncogene kinases from other protein kinases of the cell, which attach phosphate groups to the serine or threonine moieties of their targets. Recently, the *src* protein and a related protein have been found to phosphorylate certain membrane phospholipids. Proteins specified by the *ras* oncogene group have a contrasting behavior, in that they seem to be able to bind or degrade guanosine nucleotides.

Oncogenes and Growth Factors

Recent work has shown that the gene product of one retroviral oncogene, designated *sis*, is related to a subunit of platelet-derived growth factor, a protein actively involved in tissue repair. Similarly the gene product of a second retroviral oncogene, *erb*-B, represents a fragment of the cellular receptor for epidermal growth factor. These findings not only have great import for identifying the nature and mechanism of action of certain oncogenes, but also have brought together two exciting fields of study: oncogenes and growth factors. The work suggests that the intercellular signals for regulating growth (the growth factors) are intimately connected with intracellular mechanisms of growth regulation (oncogene proteins).

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

MAJOR PROBLEMS IN THE FIELD

How Oncogenes Work

Having learned to detect oncogenes and their encoded proteins, researchers now are faced with the more difficult task of discovering how individual oncogenes work to cause cancer. It is assumed that the proteins encoded by oncogenes affect the normal cellular pathways that govern growth. How do they do this?

The *src* and *ras* proteins are representatives of the large class of oncogene proteins that are found in the cytoplasm of the cell or near the inner surface of the cell membrane. Other proteins, specified by the *myb* and *myc* oncogenes, are found in the nucleus, where they may control expression of certain cellular genes. The enzymatic functions of these various proteins can provide clues to their physiological roles. Of related interest are cellular regulatory processes in which simple actions of enzymes can result in large signal amplification and produce diverse responses in the cell. Presumably, each of the oncogene proteins will eventually be found to have an enzymatic or regulatory activity that allows it to perturb a cellular growth-regulating pathway. But for the moment, we are ignorant of the nature of these growth-regulating circuits and pathways and of the mechanisms by which relatively simple proteins can elicit the multiplicity of traits associated with cancer cells. This problem of oncogene protein function represents the major challenge of the field for the coming decade.

Interaction of Oncogenes with the Cell Cycle

The simplest view of the growth status of the cell would portray two states—growing and nongrowing. In reality, a complex series of events occurs when a cell emerges from a quiescent state and enters a state of exponential growth. This latter state is characterized by a well-ordered series of events, termed the "cell cycle," that includes chromosomal replication and mitosis. It is probable that some of the complex regulatory decisions required during the cell cycle involve proto-oncogenes. For example, a rapid increase in expression of the normal *myc* gene is one of the initial events that occurs during emergence from quiescence.

Cellular Differentiation

Cancer is a disorder of differentiation as well as of cell proliferation. Stem cells in many tissues form a pool of proliferating cells, some of whose progeny leave the pool and evolve (i.e., differentiate) into specialized cells that constitute the functioning tissue. Stem cells display immortality; they can, in principle, multiply without limitation throughout the lifetime of the organism. Their differentiated daughter cells, however, lose their ability to divide indefinitely. One mechanism by which cancer cells can achieve unlimited replicative potential may depend on inhibiting the ability of these cells to differentiate, trapping them in the pool of replicating stem cells.

One possible role of oncogenes is to maintain the tumor cell in an undifferentiated state, protecting it from progression to the differentiated state with concomitant loss of replicative ability. The process of differentiation and its regulation by (proto-) oncogenes represent major problems in cell biology.

Which Cancer Cell Phenotypes Are Induced by Oncogenes?

The most commonly used definition of an oncogene implies an ability, on the part of a gene, to confer unrestrained, abnormal growth capacities on previously normal cells. Such a definition would suggest that many of the genes involved in the control of normal cell proliferation have the potential, following appropriate activation, to become oncogenes.

However, cancer is more than a problem of

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

unrestrained growth. Cancer cells acquire a number of other traits during their long progression from normality to malignancy. One important trait concerns the ability of these cells to evade being killed by components of the immune system that are responsible for detecting and eliminating abnormal cells that may arise in the body. A second attribute of some cancer cells, metastasis, confers an ability of these cells to invade adjacent tissues and to develop into secondary tumors at distant sites.

Little is known about the biochemical or molecular biologic mechanisms governing these important tumor cell traits. It is not even clear that specific activated genes are responsible for mediating metastasis or immune evasion. We suspect that such genes exist, and that they may properly be included under the rubric of oncogenes, because they probably make important contributions to tumorigenesis. A major problem concerns the nature of these genes and how they are able to produce the variety of traits associated with cancer cells.

Participation of Oncogenes in the Multiple Steps of Carcinogenesis

Carcinogenesis is believed to involve a series of independent, sequentially occurring steps that enable a cell and its descendants to evolve toward malignant growth. Certain carcinogens, termed "initiators," trigger the process of tumorigenesis. Others, called "promoters," facilitate the development of "initiated" cells into full-blown tumor cells. The effects of initiators and promoters on oncogenes are poorly understood. Several of the early steps may involve activation of various growth-promoting oncogenes. Other alterations may propel the cell toward invasion of surrounding tissue and spread to distant sites.

In one very specific case, that of rat embryo fibroblasts, at least two independently activated oncogenes are required to convert normal cells into tumor cells. Although the generality of this observation remains to be demonstrated, the existing evidence underscores the potential complexity, at the molecular level, of the tumorigenic process. The nature and number of molecular events that transpire in the creation of most human tumors are unknown.

How Many Potential Oncogenes Are Harbored by the Cellular Genome?

The cellular genome carries 50,000 or more distinct genes. Of these, only a small portion are concerned with growth control and with other traits expressed by cancer cells. The number of known proto-oncogenes found in the cellular genome now exceeds 20. These have been discovered by virtue of their association with various retroviruses, or by transfection, or by use of cloned DNA probes derived from the first two approaches. It is likely that additional oncogenes remain to be discovered, perhaps by use of new detection procedures.

How Are Oncogenes Activated From Normal Proto-Oncogenes?

An important question concerns the multiplicity of molecular mechanisms that can mediate the activation of proto-oncogenes into their oncogenic versions. For example, as already discussed, the *ras* oncogenes can be activated by discrete point mutations affecting single nucleotides. Amplified versions of *ras* genes are also found in some human malignancies. Another activated oncogene, the *erb*-B gene, results from the truncation of the gene that encodes epidermal growth factor receptor. As a third example, the *myc* gene becomes activated following chromosomal translocation. The precise mechanisms or agents (such as chemical mutagens) that may lead to these changes must be explored. Although much progress has been made in studying environmental carcinogens and their

Copyright © National Academy of Sciences. All rights reserved.

mode of activation within cells, it is still not clear precisely how many of these activated compounds are able to create oncogenes or induce human cancers.

What Might Reverse or Inhibit Oncogenes?

The panel believes that oncogenes lie at the heart of the cancer process and that they represent the centrally acting agents of cellular transformation, rather than peripheral secondary or tertiary consequences of this process. Much oncogene research has been directed toward understanding the causative mechanisms of cancer, but the discovery of these genes also provokes interest in eventual development of anticancer therapeutics. Reversal of the action of oncogenes or their encoded proteins may allow a reconversion of tumor cells to normal cells. Because certain oncogene proteins are structurally different from their normal counterparts, agents might be developed that specifically recognize and antagonize these proteins.

A major challenge is the development of pharmacologic antagonists of oncogene action. Such antagonists would provide an alternative to the cytotoxic agents currently used in cancer therapy. An alternative therapy might depend upon immunologic reagents (e.g., antibodies), which would recognize and destroy cells carrying oncogene proteins. Unfortunately, such an approach may be extremely difficult because many oncogene proteins are located deep within the cell and thus are hidden from immune mechanisms that are able to survey only the cell surface.

An important, relatively unexplored, possibility concerns the natural mechanisms that the organism employs to check cellular growth. Just as there are growth-stimulating factors, there may be corresponding growth-retarding factors. The expression of proto-oncogenes may normally be counterbalanced by growth-antagonizing "anti-oncogenes," which would be valuable in cancer prevention or treatment.

What Types of Genes Confer Hereditary Susceptibility to Cancer, and How Are These Genes Related to Oncogenes?

Many people have inborn susceptibilities to certain types of cancer. In some cases, the genetic factors conferring susceptibility exert their effects by depriving the individual of normal immune defenses for protecting DNA from accumulating large numbers of mutations. But in most cases, the susceptibility is less general, predisposing to specific types of cancer.

The bases for these susceptibilities are not known, but their genetic components are different from the oncogenes described here. The susceptibility genes are associated with an individual from the time of fertilization of the egg; oncogenes are not genetically transmitted in egg or sperm but develop instead in specific tissues during an individual's lifetime. Several of the susceptibility genes have been associated with specific chromosomal aberrations. The identity of these genes and the nature of their functions are obscure at present.

POTENTIAL AREAS OF RAPID DISCOVERY

GROWTH FACTORS AND SECOND MESSENGERS

The growth of most normal cells in the body seems dependent on growth factors, hormones that are transmitted from cell to cell and provide the external stimuli that trigger growth. Cancer cells often can grow in the absence of the normally required exogenous growth factors. From this perspective, oncogenes can be viewed as agents that confer on cells a growth pattern that is independent of external growth factors.

For example, one oncogene has been found to cause the cell itself to make inappropriately large amounts of a growth factor so that the cell's growth becomes independent of exogenous factors. Another oncogene

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

encodes the structure of an altered growth factor receptor in the cell; this altered receptor may make the cell behave as if growth factors are present even when they are not.

These data and hypotheses have resulted in a merging of two previously separate fields of cancer cell biology, oncogenes and growth factors. An understanding of the mechanism of action of growth factors and their receptors will be central to understanding the physiology of the cancer cell.

Ion Channels and Second Messengers

Regulation of cell growth has often been studied by depriving normal cells of growth stimulatory factors and then inducing them to grow by exposure to these factors. Within seconds of exposure to growth stimulatory factors, a variety of changes are seen in the cell. To cite some examples, Na^+ and H^+ are exchanged across the outer membrane of the cell, the intracellular pH rises, free calcium in the cell rises dramatically, and several of the phospholipids of the membrane are degraded to new compounds.

These phenomena illustrate the activity of so-called "second messengers"— compounds in the cell whose changing concentrations transmit important signals from one cellular constituent to many others. It is highly likely that oncogene proteins and their normal versions modulate the concentrations of these second messengers, which then convey growth regulatory information to various parts of the cell. (It may be that some of the ion-regulating mechanisms that originally evolved for mediating cellular growth control were subsequently adapted for modulating ionic states of nerve cells during conductance of excitatory signals.)

These various considerations indicate that two areas of rapid development over the next several years involve study of intracellular second messengers and study of proteins that regulate movement of ions across membranes.

Biochemical and Structural Analysis of Oncogene Proteins

Study of oncogene proteins has been hampered because only small amounts of material were available. Successful techniques for cloning the oncogenes and expressing their proteins in bacteria have made increased amounts of these proteins available for study. It soon will be possible to conduct structural analyses of these proteins by means of directed modification of the proteins themselves and x-ray crystallographic analysis. Moreover, the recently developed techniques of directed DNA mutagenesis (whereby specific nucleotides can be altered) will make it possible to alter oncogene sequences and thereby redesign specific portions of the proteins they encode.

Use of synthetic oligopeptide immunogens and monoclonal antibodies will enormously facilitate the localization of these proteins to specific cellular sites, the identification of other proteins with which these proteins interact, and the characterization of complex, multicomponent reactions in which these proteins participate.

Genetic Analysis in Yeast and Fruit Flies

Genetic analysis in well-studied lower organisms provides a powerful means of looking at properties of proto-oncogenes and oncogenes and their role in cellular and organismic physiology.

In yeast, *Saccharomyces*, and to a lesser extent in the fruit fly, *Drosophila melanogaster*, it is possible to inactivate a proto-oncogene genetically and to study subsequent effects on the organism's viablity, growth, and development. Identification of other genes that interact with oncogenes will provide a new dimension to the current understanding of cellular pathways that are perturbed by oncogene action. Such genetic manipulation is impossible with mammalian cells. Implicit in these experiments is the notion that the

Copyright © National Academy of Sciences. All rights reserved.

molecular mechanisms governing cellular replication in yeast and flies are similar to those in mammalian cells.

The oncogenes discussed until now encode protein products that induce the cell to proliferate. The action of these oncogenes may be balanced by other genes that function to inhibit growth. If such antigrowth genes exist, inactivation of these genes would physiologically mimic the action of oncogenes that actively promote growth.

Some data suggest that certain genes that confer a hereditary, familial predisposition to cancer may act by failing to inhibit growth. Thus, individuals born with an inadequate complement of these "antigrowth" genes may have a lifelong risk of increased cancer incidence. How can we look for these "antigrowth" genes or, more precisely, for the absence of them? It should be possible to use genetics to identify growth-restraining genes in yeast and to employ these yeast genes to find counterparts in mammalian cells.

A further use of genetics derives from our newly gained ability to introduce genes into the germ lines of *Drosophila* and mice. Studies in which oncogenes are introduced into these germ lines will allow researchers to look at the effect that these genes have on normal development of the embryo or on specific organs within the embryo. Such studies promise to provide important new insights into the roles of (proto-) oncogenes in the development of complex organisms and the interaction of certain oncogenes with specific tissues.

The Search for New Oncogenes

Discovery of additional oncogenes is limited by the techniques currently used to detect them. New assays for oncogene activity will be developed over the next several years and presumably will reveal oncogenes whose presence in human DNA was not apparent using existing techniques.

Certain types of oncogenes should be discovered by virtue of their ability to transform indicator cells other than the currently used NIH3T3 mouse fibroblasts. Other oncogenes may be found following direct manipulation of the normal human DNA. Yet others may be found by virtue of their association with specific chromosomal rearrangements.

Cancer Diagnosis

Several reagents soon will be available to detect oncogenes and distinguish their products from those of normal counterpart genes. These reagents may allow earlier identification and greater specificity in detecting tumors than is currently possible.

The success of such diagnosis is predicated upon an issue as yet unresolved: How often are altered oncogenes involved in the causation of human tumors? About 20 percent of human tumor DNAs yield active oncogenes upon transfection into NIH3T3 cells. However, this assay has limitations. In fact, we already know that some oncogenes such as *myc,* which was identified because of its involvement in chromosomal translocations, do not affect NIH3T3 cells and would have been overlooked in such an assay. In other cases, activation of a proto-oncogene may be a necessary but not sufficient step in carcinogenesis and may not warrant a diagnosis of an actively growing cancer.

As methods are improved for discovering new oncogenes, it is the panel's opinion that all human tumors eventually will be found to carry cellular oncogenes. Because oncogenes are important determinants of cellular phenotype, we expect that oncogene diagnosis will provide important prognostic indicators for the clinician dealing with specific tumors. Oncogene diagnosis also may permit better delineation among cancer types and provide better clues to external triggering agents.

NEED FOR CONTINUED BASIC RESEARCH

The great amount of progress in oncogene studies has been possible because of the

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

basic biologic research that preceded and led to the discovery of oncogenes. Continued study in fields such as membrane biology, signal transduction, cellular differentiation, molecular genetics, and cellular immunology will come to bear on the problem of cellular transformation and tumorigenesis. Conversely, the lessons learned from oncogene research will have important implications for many of the unsolved problems of modern biology.

Although the study of oncogenes cannot be expected to resolve all the diverse phenomena that constitute the biology of cancer, we believe that the progress to date has already placed us close to a clear understanding of many of its central mechanisms. Further progress will depend on continuing support of the basic biologic disciplines whose successes have proved so useful in confronting a multitude of biologic problems, cancer being only one of them.

SUMMARY

Within the past 5 years, molecular and cell biologists have made great advances in identifying the molecular mechanisms of carcinogenesis. They have found specific genes within tumor cells, termed "oncogenes," which are responsible for creating many of the abnormal traits of cancer cells. These discoveries provide the foundation for a complete understanding of the molecular basis of cancer within the coming decades.

While these genes have been assigned important roles in carcinogenesis, their precise mechanism of action remains unclear. Much work will be pursued on the structure and enzymatic functions of the oncogene proteins. Thus, certain oncogenes may affect the levels of important compounds in the cell that regulate growth. Others may regulate the expression of certain cellular genes. Yet others may regulate the expression of secreted growth-regulating hormones or receptors used by the cell to detect these hormones.

Future advances in understanding oncogene function will depend on exploiting a variety of experimental tools made available by such diverse basic biologic disciplines as pharmacology, molecular biology, genetics, and cell biology. The presence of related genes in organisms such as fruit flies and yeast will make it possible to study the oncogenes in the context of these genetically simpler organisms. Future advances also will focus on how chemical carcinogens can activate oncogenes and how other agents may antagonize oncogene function and the growth of the cancer cell. New genes may be found that function to control oncogenes and inhibit their functions. Such advances may lead over the next decades to insights into novel therapies that interfere specifically with the growth of cancer cells.

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Report of the Research Briefing Panel on Interactions Between Blood and Blood Vessels (Including the Biology of Atherosclerosis)

Copyright © National Academy of Sciences. All rights reserved.

Research Briefing Panel on Interactions Between Blood and Blood Vessels (Including the Biology of Atherosclerosis)

Alfred P. Fishman (*Chairman*), Director, Cardiovascular-Pulmonary Division, Department of Medicine, Hospital of the University of Pennsylvania, Philadelphia, Pa.

Ramzi S. Cotran, Professor of Pathology, Harvard Medical School, and Pathologist-in-Chief, Brigham and Women's Hospital, Boston, Mass.

Harriet P. Dustan, Director, Cardiovascular Research and Training Center, University of Alabama in Birmingham, Birmingham, Ala.

Howard A. Eder, Professor of Medicine, Albert Einstein College of Medicine, Bronx, N.Y.

Judah Folkman, Professor of Pediatric Surgery and Professor of Anatomy, Harvard Medical School, and Senior Associate in Surgery, Children's Hospital, Boston, Mass.

Robert Furchgott, Professor of Pharmacology, Downstate Medical Center, Brooklyn, N.Y.

John C. Hoak, Professor and Chairman, Department of Medicine, University of Vermont, Burlington, Vt.

Eugene Renkin, Professor and Chairman, Department of Human Physiology, School of Medicine, University of California, Davis, Calif.

John Repine, Professor of Medicine and Pediatrics, Webb-Waring Lung Institute, Denver, Colo.

Russell Ross, Professor and Chairman, Department of Pathology, School of Medicine, University of Washington, Seattle, Wash.

Una Ryan, Professor of Medicine, University of Miami School of Medicine, Miami, Fla.

Alan D. Schreiber, Professor of Medicine, Hospital of the University of Pennsylvania, Philadelphia, Pa.

Thomas W. Smith, Chief, Cardiovascular Division, Brigham and Women's Hospital, and Professor of Medicine, Harvard Medical School, Boston, Mass.

Peter Ward, Interim Dean, University of Michigan Medical School, Ann Arbor, Mich.

Albert Winegrad, Professor of Medicine, Cox Institute, Philadelphia, Pa.

Staff

Barbara Filner, *Director*, Division of Health Sciences Policy, Institute of Medicine

Barbara Mandula, *Consultant*

Naomi Hudson, *Administrative Secretary*

Copyright © National Academy of Sciences. All rights reserved.

Report of the Research Briefing Panel on Interactions Between Blood and Blood Vessels (Including the Biology of Atherosclerosis)

In atherosclerosis (a form of hardening of the arteries), blood constituents are deposited in the blood vessel wall, causing clotting and inflammation. If the vessel becomes blocked, the consequences can be extremely serious. Medical care and medical research have long been concerned with the complications of atherosclerosis, ranging from heart attacks and strokes to peripheral vascular insufficiency and kidney disease. Complications of atherosclerosis are relatively common, account for an extraordinary toll of chronic impairment and early deaths, and result in staggering economic costs. Despite a 30 percent decline in death rates from heart attacks and strokes in the United States in the past 15 years, complications of atherosclerosis still cause almost half of all deaths in this country. The disease was identified as the leading cause of days lost from work in 1977 and was estimated to have cost almost $40 billion in health expenditures and lost productivity in the United States.

It is now evident that atherogenesis, the disease process that culminates in atherosclerosis, entails a complex interplay between blood and the vessel walls that contain it. In recent years, great progress has been made in understanding this interplay, and this should enhance our ability to treat and prevent the diseases that result from abnormal blood/vessel interactions. The new knowledge has resulted from interdisciplinary efforts drawing on fresh concepts and new techniques of molecular and cell biology, physics, chemistry, and genetics.

Atherogenesis has been recognized as just one example, albeit one of overriding practical importance, of the disease processes that occur when there is a breakdown of normal interactions between blood constituents and the vessel wall. Studies of such interactions also pull together information about normal and abnormal functions of the circulation, the lungs, and the heart. These studies have the potential to provide useful information for a wide variety of human ills.

This paper considers first the characteristics of the blood vessel wall, especially the endothelium, which is in contact with the blood. Second, the paper discusses the properties of some components of the blood that interact with vessel walls. Third, examples are provided of interactions between blood components and the vessel wall that are of broad biomedical significance. Each of these examples is intended to illustrate normal interplays and the consequences of their

Copyright © National Academy of Sciences. All rights reserved.

disordering by disease or injury. Atherogenesis will be discussed in some detail in the context of recent findings. Finally, opportunities that require further investigation are identified.

THE CIRCULATORY SYSTEM

Each component of the circulatory system has its own distinctive character, metabolism, behavior, and disorders. Blood flowing through this system consists of a clear liquid plasma in which distinctive cells are suspended. The large vessels serve as conduits, but the tiny vessels, principally the capillaries, serve as semipermeable barriers for exchange, between blood plasma and tissues, of oxygen, carbon dioxide, nutrients, wastes, and other substances. Capillaries also have functions peculiar to the tissue or organ in which they are situated. In the lungs, for example, the vast capillary network acts as a physical filter to prevent particles, such as clots, from reaching and harming vital organs, particularly the brain and heart.

Large arteries are subject to atherosclerosis and both arteries and veins incur thrombosis, or clotting. Capillaries can be damaged by inflammatory diseases, such as occurs in the kidney in glomerulonephritis.

Although the architecture and basic structures of the circulatory system have been known for a long time, the detailed molecular mechanisms by which the system carries out its functions are only now being unraveled. The use of modern biologic techniques has made it possible to analyze interactions among the various components of the system and to accrue new information about control mechanisms in health and disease.

ENDOTHELIUM

The inner lining of blood vessels is a single layer of cells, the endothelium, which has remarkable properties. It has an apparently simple structure and once was thought to fulfill fairly simple tasks; now it is known to have biologic functions that are many, complex, and sometimes unique to the organs in which the blood vessels are located. In general, the monolayer serves as a selective barrier to the movement of fluids and cells between blood and tissue, as an antithrombotic lining, and as a metabolic organ in its own right.

The lining of vessels normally allows the passage of blood for a lifetime without clotting and without its constituents becoming denatured. No artificial substitute has yet been created that can do that. But an injury upsets the dynamic balance between the vessel wall and blood constituents. A subtle injury, in the form of sustained high blood pressure or high blood cholesterol levels, appears to enhance the permeability of the vascular wall to plasma constituents and enables cells to migrate from the blood stream and possibly contribute to the beginnings of disease.

Information transfer among the components of the vessel wall, and between the endothelial lining and blood constituents, is prerequisite for normal structure and function of the circulatory system. In capillaries, whose wall consists entirely of endothelium, communication between contiguous cells somehow ensures that the vessel lining is a continuous monolayer. In arteries, whose walls are more complex, endothelium communicates with muscle cells that form the next layer outside of it, generally influencing the state of contraction or dilation of the vessel. How this cell-to-cell communication operates is of great investigative interest, and the results of current studies may radically change traditional concepts of how blood flow and pressure are controlled.

Endothelium as a Metabolic Organ

New techniques of culturing endothelial cells have made it possible to study their biochemical activity under controlled conditions and to harvest sufficient material for biochemical study. A combination of in vivo and in vitro studies, for example, has disclosed

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

the vital role of pulmonary capillary endothelium in maintaining systemic blood pressure. On the luminal surface of pulmonary capillary endothelium, the surface in contact with blood, angiotensin-converting enzyme degrades bradykinin (a vasodilator) to inactive products. The same enzyme also converts angiotensin I to angiotensin II (a vasoconstrictor), thereby further helping to sustain systemic blood pressure. In addition, pulmonary capillary endothelium generates angiotensin III, a hormone that acts on the kidneys to influence blood pressure by affecting salt and water balance.

Endothelial cells also participate in maintaining the fluidity of the blood while ensuring that clot formation occurs locally as needed to limit bleeding. Sequences of enzymatic reactions ("cascades") have been identified that either form clots or dissolve clots. Cellular components of the blood, particularly the platelets, also contain powerful substances that lead to clot formation when they encounter injured endothelial lining. The endothelial lining locally generates a hormone (prostacyclin) that modulates the clotting response. Bradykinin, released and destroyed locally, enhances the generation of prostacyclin. These are complicated interplays, regulating the coagulation-anticoagulation systems. Much remains to be learned about the coordinated responses, the mechanisms that prevent under- and overreaction during injury or stress, and the imbalances that arise from kidney failure, or a variety of prescribed drugs, or inherited disorders involving these components.

The growing understanding of endothelial functions has been paralleled by insights into its structural complexities. Ultrastructural specialization of the endothelial cell surface occurs, allowing it to localize diverse functions and carry them out more effectively. Some domains of functional activity are distinctive because of electrostatic charge, while others occur as invaginations in which enzymes are concentrated. How the various molecular activities are organized and orchestrated is currently being explored in depth.

Further study is needed to compare the behavior of endothelial cells in vivo and in vitro, to determine optimal methods for growing endothelial cells from different sites and organs in vitro so as to simulate natural conditions, and to compare endothelial function in large and small vessels. Endothelial cells in culture also are beginning to be used to produce biologically useful substances, such as plasminogen activator, which is important in clot dissolution, and to test pharmacologic agents for toxicity. This research also should continue to be productive.

Transport of Material Across Endothelial Cells

Substances must cross the endothelium in order to leave the blood and be delivered to tissues. Scientists are using a variety of techniques to explore the mechanisms involved in the passage of substances between blood and tissues and the reasons for the different structures of endothelium in different parts of the body.

Endothelium that lines the capillaries of muscle and lungs is continuous; the cells of the monolayer are linked by tight junctions. In contrast, endothelium that lines liver sinusoids and kidney glomerular capillaries has gaps between adjacent endothelial cells so that the adjacent basement membrane, rather than the endothelial layer, constitutes the initial barrier between blood and tissue.

Until recently, the passage of large molecules, such as proteins, was explained in terms of a "pore theory," which envisioned movement through holes or tunnels in the capillary wall. However, experimental observations have made this theory less convincing. At present, there is particular interest in the biochemical attributes of the endothelial cell and the underlying interstitial space as determinants of the transcapillary passage of blood constituents. In addition, attempts continue to relate anatomical structures, such as vesi-

Copyright © National Academy of Sciences. All rights reserved.

cles, to the transcapillary passage of charged and uncharged particles. The rapidly developing field of transendothelial transport has added fresh impetus and direction to these studies.

Communication Between Endothelium and Smooth Muscle

The state of contraction of the smooth muscle in the walls of small arteries and arterioles largely determines blood pressure and is responsible for adequately distributing blood to all parts of the body. A variety of natural regulatory mechanisms automatically adjust the state of contraction, or vascular tone. Some pharmacologic agents act directly to increase or decrease vascular tone, that is, to cause vasoconstriction or vasodilation.

Only recently has it been demonstrated that certain agents act as vasodilators by acting on the endothelium rather than on the muscle of the vessel wall. It has been proposed that a humoral substance, "endothelium-derived relaxing factor" (EDRF), mediates this dilation. Biochemical substances that seem to exert a vasodilating effect through EDRF include acetylcholine, substance P, adenosine diphosphate, and adenosine triphosphate. The finding has important implications for local regulation of blood supply both under normal conditions and under conditions of vascular disease in which physical and chemical links between cells of the vessel wall are often markedly deranged.

The chemical identity of EDRF has not yet been established. It is possible that EDRF has a central role in maintaining proper blood pressure and that further study of it will help in clarifying and controlling abnormalities in blood pressure.

Angiogenesis

The growth of new capillaries is greatly dependent on the activity of endothelial cells. The endothelial cells that line capillaries are usually in a resting state. Their rate of proliferation is so low that turnover times are measured in years. However, these endothelial cells are capable of sudden growth spurts during periods when new capillaries are rapidly forming, for example, during wound healing. The process of new capillary formation is called "angiogenesis," and cell turnover times during angiogenesis are measured in days.

Protracted angiogenesis is observed in such diverse pathologic states as arthritis, psoriasis, diabetic retinopathy, and chronic inflammation; the most sustained and intense angiogenesis is associated with solid tumors. In many of these abnormalities, angiogenesis itself contributes to the disease process. In arthritis, new capillaries can invade and destroy joint cartilage. In diabetes, new capillaries in the retina of the eye hemorrhage and cause blindness. Progressive growth of tumors and their ability to form metastases depend upon continuous angiogenesis.

Because of techniques introduced since 1970, it is now possible to investigate angiogenesis experimentally. Capillary endothelial cells have been cloned and grown in vitro, where they form tubes, branches, and whole capillary networks. Special polymers have been developed for the sustained release of angiogenic factors in vivo. Bioassays for angiogenic activity have been perfected in the chick embryo and in the rabbit and mouse cornea.

Application of these techniques is revealing some components of the angiogenic process. As an early step in capillary formation, endothelial cells release specific enzymes that punch holes in the vascular basement membrane. Endothelial cells migrate through these holes. The cells align in a linear configuration to form a new sprout. The sprout elongates through interstitial tissue toward an angiogenic stimulus such as a small tumor. Sprout elongation depends upon chemotaxis of endothelial cells at the tip of the sprout, division of cells farther back, and production

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

of interstitial collagenases, probably by all of the endothelial cells in the sprout.

Techniques developed primarily for the study of tumor angiogenesis have recently been used to explore the mechanisms of angiogenesis associated with processes such as wound healing. Endothelial cells and debris-scavenging macrophages play major roles. In the depth of a fresh wound, tissue oxygen levels are extremely low because capillaries have been damaged and blood flow interrupted. In a setting that is poor in oxygen, macrophages that migrate into the wound secrete angiogenesis factors. As new capillaries are attracted into the wound, oxygen tension rises; the macrophages then reduce or shut off their stimulation of angiogenesis. The details of this feedback mechanism remain to be worked out, and the macrophage angiogenic factor has not yet been purified.

Blood Constituents

Blood is composed of circulating cells suspended in a liquid, the plasma. Each component of blood has access to endothelium and virtually all bear electrostatic charges that interact with each other and with the surface charges on the endothelial lining. The composition of the blood is automatically adjusted to the various needs of the body, including respiratory gas exchange, maintenance of blood pressure, and maintenance of fluid and electrolyte balance. Blood constituents interact with each other as well as with the vessel wall. A few of the blood constituents that have crucial roles in interacting with the vessel wall are discussed below.

Platelets

Platelets are essential elements in blood clotting and also have a central role in the process of atherogenesis. Upon contact with collagen or macrophages in an area of vascular injury, platelets aggregate and their storage granules discharge biologically active substances that promote the formation and organization of a clot to plug the breach in the endothelial lining. Enzymes and growth factors released by the aggregated platelets attract other platelets, evoke spasm and proliferation of adjacent vascular smooth muscle, and draw coagulation proteins to the area for the assembly of the clot. These reactions are controlled by molecular feedback mechanisms. For instance, the activity of thromboxane, released by platelets to promote local vascular constriction and accumulation of other platelets, is modulated in its clot-promoting effects by prostacyclin (an anti-aggregating and vasodilator agent) liberated by endothelium.

New techniques have greatly expanded our knowledge of platelet structure and function. The surface of the platelet displays specialized molecular organization, which includes receptors that provide binding sites for macromolecules that engage in complex immunologic reactions, and other receptors that have to do with the clotting process and that bind fibrinogen and von Willebrand factor. These receptors constitute ultrastructural foci for the orderly assembly of the clot. As the coagulation proteins fall into place on the platelet surface, the clotting process is dramatically accelerated by a surge in formation of thrombin.

There are still large gaps in the understanding of how platelets engage in their complex interactions with the vessel wall, plasma factors, and other blood cells. Their role in the inflammatory process, for example, remains enigmatic. Platelets, in addition, are a storehouse for powerful enzymes that differ strikingly in structure and function. How they interrelate in response to injury requires much further study.

Leukocytes

Leukocytes, or white blood cells, are part of the body's defense system. They are drawn chemotactically to areas of damage or invasion by foreign agents and participate in the

Copyright © National Academy of Sciences. All rights reserved.

ingestion, walling off, and elimination of the offending agent. Upon stimulation by appropriate chemicals, leukocytes generate oxygen radicals and release substances capable of destroying microorganisms. We are now beginning to understand how the body protects itself against inappropriate activity of these powerful chemicals. Phagocytic cells involved in the inflammatory response, as well as other cells of the body, are equipped with intracellular antioxidant enzymes, such as catalase and glutathione peroxidase; in addition, there are circulating neutralizers, such as ceruloplasmin and alpha-1 antitrypsin, to prevent these chemicals from damaging the host when they are released. It is anticipated that awareness of how leukocytes inflict their damage will make it possible to develop medications to protect people against their toxic products.

Plasma Lipoproteins

Because lipids are insoluble in an aqueous medium, they are transported in plasma as macromolecular complexes called lipoproteins. Of the various types of lipoproteins, low-density lipoprotein (LDL) carries one-half to two-thirds of all the cholesterol in plasma. LDL has cholesterol ester as its core lipid and unesterified cholesterol on its surface. High levels of LDL-cholesterol predispose to atherosclerosis. But another form of lipoprotein, high-density lipoprotein (HDL), may be anti-atherogenic.

Cells of many different tissues have specific receptors for LDL and transport it into the cytoplasm, where it subsequently is degraded in lysosomes. In normal circumstances, however, the entry of LDL into the artery wall is predominantly receptor-independent and occurs by a nonspecific process known as transcytosis. Reverse transport of LDL also occurs and avoids undue accumulation of cholesterol within the cell. The mechanisms involved in reverse transport and in the feedback control of intracellular cholesterol concentrations are under investigation, as is the mechanism of LDL transport into and out of endothelial cells following injury to the vascular lining.

Macrophages and endothelial cells have specific receptors for LDL, although their rate of uptake of LDL is low. Macrophages also have specific receptors for LDL that has been chemically modified or altered by contact with endothelial cells. The rate of uptake of modified LDL by this receptor is considerably higher than that of unmodified LDL, and foam cells (macrophages filled with cholesterol) can be readily produced by interaction of modified LDL with macrophages. LDL also apparently alters endothelial cell membranes in a way that allows access of lipoproteins into the vessel wall. When LDL is present in the vessel wall, it may cause proliferation of smooth muscle cells, which migrate into the lining of the blood vessel of the affected artery.

Considerably less is known about the interaction of vessels with the larger lipid-rich lipoproteins, such as chylomicrons. These are triglyceride-rich particles that appear in the circulation after the ingestion of fat and in certain disease states. These are removed by the liver, which then secretes the cholesterol in the particles in the form of VLDL, which is metabolized in part to LDL. It is by this mechanism that dietary cholesterol eventually enters the vessel wall.

In persons with a disorder called Type III hyperlipoproteinemia, atherosclerosis occurs frequently. In these patients, the chylomicron remnants and VLDL remnants enriched in cholesterol accumulate in the plasma in a form called beta-VLDL. Macrophages from patients with Type III hyperlipoproteinemia have receptors for beta-VLDL, which probably accounts for the accumulation of foam cells in vessel walls in the distinctive skin lesions of these patients.

Earlier studies had suggested that high-density lipoproteins (HDL) are involved in transporting cholesterol out of macrophages. HDL contains a number of apoproteins (the protein portion) of lipoproteins. The major one in HDL is apo A-I. A specific receptor for apo

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

A-I has been found on endothelial cells, on vascular smooth muscle cells, and on some liver cells. It has been demonstrated recently that the apo A-I receptor on nonliver cells is regulated by the cholesterol content of the cells. Thus, HDL may transport intracellular cholesterol out of the vessel, becoming enriched in cholesterol while doing so. How the receptor for HDL on liver cells is regulated is not yet known.

Mechanisms of entry of LDL into the vessel wall and handling of LDL when endothelial cells are damaged require further investigation. The extent of interaction of chylomicron remnants or VLDL remnants with the vessel wall also needs study. An exciting new area of investigation involves the mechanism by which HDL, which can bind to specific receptors on vascular endothelial cells and on smooth muscle cells, transports cholesterol from these tissues. New tech niques, such as production of monoclonal antibodies for the various apolipoproteins, will be important for studying the metabolism of apolipoproteins and their role in atherogenesis.

EXAMPLES OF BIOMEDICALLY IMPORTANT INTERACTIONS

Molecular interactions between the constituents of blood and the endothelium operate continuously to permit normal growth and development and to sustain the metabolic activities of daily life. Respiratory gases, water, nutrients, and other molecules must move between endothelium, plasma, and the cells of the blood. Electrolytes pass back and forth between blood and tissues where they trigger and drive metabolic machinery. Hormones travel long distances in blood to activate and participate in biological processes far removed from their sites of origin. Each of these systems is governed by elaborate control mechanisms that continuously monitor and restore dynamic equilibria. To illustrate the operation, complexity, and control of the biologic systems and their regulatory mechanisms, three examples will be considered: blood clotting, the inflammatory process, and atherogenesis.

Blood Clotting (Thrombosis)

The tendency for blood to clot when a blood vessel is severed is a safeguard against bleeding to death. In living systems, this tendency is counterbalanced by mechanisms to ensure blood fluidity; inappropriate or excess clotting may lead to serious clinical syndromes. Thrombosis plays a critical role in initiating, propagating, and occluding the atherosclerotic lesion, thereby constituting the precipitating cause of acute myocardial infarction, stroke, and peripheral vascular disease. Thrombosis is a precursor of thromboembolic disease, a common and life-threatening complication of surgery, heart failure, and cardiac arrhythmias. Improved understanding of thrombotic events—involving endothelium and blood constituents—holds the prospect of better therapeutic and prophylactic interventions.

The complex sets of reactions involved in clot formation and removal have been examined experimentally for years, but more pieces of the puzzle continue to be found. For example, one recent discovery is Protein C, a vitamin-K-dependent coagulation factor and a potent anticoagulant and fibrinolytic agent. Its concentration is regulated by an elaborate biochemical control system: Protein C is activated by thrombin bound to the protein thrombomodulin on endothelium; activated Protein C is neutralized by an inhibitor present in the plasma. Protein C appears to be crucial for life and several infants who lacked Protein C because of a genetic defect died of generalized thrombosis shortly after birth.

The Inflammatory Process

The inflammatory process is a vital component of the defense mechanisms of the body. A cellular response, abetted by an increase in endothelial permeability evoked

Copyright © National Academy of Sciences. All rights reserved.

by inflammatory peptides and lipids, contributes to the leakage of water, salts, and proteins from plasma into the extravascular environment. Inherent in this process is the migration of leukocytes from the blood to the site of injury. En route, these blood cells traverse endothelium.

New concepts and techniques are currently being applied to the interplay between the blood and endothelium in the inflammatory process. Promising lines of research include the effects of particular types of endothelial injury on the inflammatory response, the response of endothelium to substances, such as bradykinin, that are involved in the inflammatory response, and the mechanisms that promote adherence between neutrophils (a type of white blood cell) and the endothelial cell.

Atherogenesis

Atherosclerosis is a special type of arteriosclerosis that affects large arteries, such as those of the coronary and cerebral circulations. Atherogenesis is the process that culminates in atherosclerosis. It results from misdirections and overresponses in normal biologic interplay at the molecular level. In its early stages—those most relevant to the understanding of initiating mechanisms and the institution of preventive measures—the inner layer of an affected artery is thickened by a deposition of fat, particularly cholesterol, and by an increase in the number of cells in the lining of the vessel. Two hypotheses concerning the etiology and pathogenesis of atherosclerosis continue to be studied intensively: the lipid infiltration hypothesis and the response-to-injury hypothesis. Each of these has stimulated new knowledge about the interplay between the vascular wall and the blood in atherogenesis.

Lipoproteins in Atherosclerosis

As noted above, the importance of lipoproteins as the cholesterol-carrying substances in blood has been clarified during the last decade. The recent decrease in mortality from heart attacks and strokes has been attributed, in part, to improved understanding of the lipoproteins and their relation to risk factors, such as high blood pressure, smoking, diabetes, and high levels of blood cholesterol. However, still to be resolved is how risk factors act as injurious agents and in what ways the responses to these diverse types of injury differ. New approaches to slowing, arresting, and reversing the atherosclerotic process are being sought. Among the more promising is modification of diet, coupled as needed with the use of medications designed to lower blood levels of cholesterol. One recent lead is the possible effect of certain fats present in seafood in lowering blood levels of cholesterol and low-density lipoproteins.

Response-to-Injury Hypothesis

This hypothesis for atherogenesis, proposed in 1973, suggested that the various risk factors associated with atherosclerosis somehow lead to alterations in the endothelial cells, and that these represent various forms of "injury."

An important development associated with the response-to-injury hypothesis has been the discovery of growth factors that may be important in the development of the lesions of atherosclerosis. Two of these are the platelet-derived growth factor (PDGF) and the macrophage-derived growth factor. PDGF is the principal substance in whole blood serum that is responsible for proliferation of cells, such as smooth muscle and fibroblasts, in cell culture. PDGF may be associated with initiation and progression of the lesions of atherosclerosis, because experimentally induced lesions can be prevented by agents that substantially decrease circulating platelets or inhibit platelet interactions.

PDGF plays an important role both in mobilization and proliferation of cells. It has been purified to homogeneity, character-

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

ized, and sequenced. PDGF binds to a specific high-affinity receptor on the surface of susceptible cells, and promptly induces such intracellular events as tyrosine phosphorylation of several proteins, including the receptor for PDGF; phospholipase activation leading to diglyceride formation; and formation of arachidonic acid and prostaglandin. It also leads to increased binding of lipoproteins, increased endocytosis, increased protein and RNA synthesis, and then, after 24 to 36 hours, to DNA replication and cell division. Platelet-derived growth factor also is a chemotactic agent that can attract cells, such as smooth muscle, into the cavity of the affected artery.

Experiments concerning the cellular and molecular biology of endothelium (which is also a potential source of platelet-derived growth factor), smooth muscle, the monocyte/macrophage, and the platelet will be important in understanding the lesions of atherosclerosis. Studies that examine the interactions of these cells have opened new frontiers and represent vitally important areas for further research.

Thrombosis in Atherosclerosis

Advances in our knowledge of blood clotting factors, platelets, and prostaglandins have stimulated new insights into old theories about how thrombosis is involved in all stages of atherosclerosis. Studies have demonstrated that platelets, the principal cells involved in thrombosis, may play a role in the initiation of atherosclerosis. Such studies may help to control and prevent thrombosis.

IMPLICATIONS AND PROSPECTS

The concepts and techniques of cell biology that deal with structure, function, and control mechanisms at the molecular level are now available for application to the important biomedical problem of interactions between the vessel and the blood that flows through it. To date, the major focus has been on the individual components involved in this interplay. The time is now ripe for a concerted interdisciplinary effort to understand the control mechanisms that regulate the interplay in health and the derangements arising from perturbations in the system arising from injury and disease.

Favoring a concerted effort at this time is the current capability for harvesting large quantities of endothelium grown in culture. In addition, innovative cell sorters and isolation techniques make it possible to isolate manageable quantities of individual blood cells for manipulation and study. Biochemical and biophysical techniques also provide precise tools for quantifying constitutents in plasma. Finally, receptors and channels at the level of the cell and its organelles can be examined using approaches of molecular and cell biology that were previously unavailable.

The practical rewards of this fundamental research have been suggested in the sections of this paper devoted to thrombosis, inflammation, and atherosclerosis. Many other prospects can be reasonably expected to follow greater understanding of the blood and blood vessel interactions: improved prosthetic devices, such as heart valves, vascular replacements, dialyzing membranes, and artificial organs; fresh insights into the prevention and treatment of vascular occlusions and their complications; new approaches to the prevention, diagnosis, management, and reversal of atherosclerotic lesions. Study of interactions between the blood and the blood vessels provides a dramatic opportunity for progress in fundamental science and holds extraordinary promise for practical application.

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Report of the Research Briefing Panel on the Biology of Parasitism

Copyright © National Academy of Sciences. All rights reserved.

Research Briefing Panel on the Biology of Parasitism

John R. David (*Co-chairman*), John LaPorte Given Professor and Chairman, Department of Tropical Public Health, Harvard School of Public Health, and Professor of Medicine, Harvard Medical School, Boston, Mass.

Joshua Lederberg (*Co-chairman*), President, The Rockefeller University, New York, N.Y.

George Cross, Professor of Molecular Parasitology, The Rockefeller University, New York, N.Y.

Paul Englund, Professor of Biological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, Md.

Paul Gross, President and Director, Marine Biological Laboratory, Woods Hole, Mass.

Seymour J. Klebanoff, Professor of Medicine, and Head, Division of Allergy and Infectious Diseases, Department of Medicine, University of Washington, Seattle, Wash.

Adel Mahmoud, Professor of Medicine, Chief, Division of Geographic Medicine, Case Western Reserve University, University Hospital, Cleveland, Ohio

Adolfo Martinez-Palomo, Head, Section of Experimental Pathology, Centro de Investigacion y de Estudios Avanzados del Instituto Politechnico National, Mexico City, Mexico

Staff

Barbara Filner, *Director*, Division of Health Sciences Policy, Institute of Medicine

Barbara Mandula, *Consultant*

Naomi Hudson, *Administrative Secretary*

Copyright © National Academy of Sciences. All rights reserved.

Report of the Research Briefing Panel on the Biology of Parasitism

Parasites are organisms that require another organism, a "host," for survival and that usually do some harm to the host. The field of parasitology has been limited to those organisms that belong to the animal kingdom, although bacteria, viruses, and fungi are also parasites. Of concern here are the single-celled protozoans, such as cause malaria and leishmania; multicellular worms or helminths, such as schistosomes and filaria; and arthropods, which include insects. Many of the arthropods also are vectors that transmit the protozoan and helminthic parasites.

There are two major reasons for the current intense interest in the study of parasites. First, such studies will have a large impact on our understanding of basic biologic processes such as those involved in cell growth and differentiation. Many of the human parasites have multiple hosts and may have several developmental stages in each host. Each stage involves regulated transitions from one form to another. Further, parasites have evolved very specialized adaptations that enable them to infect the host and then evade the host's defense mechanisms while they survive and replicate, utilizing for these purposes the host's metabolic processes. The mechanisms by which parasites adapt to their environment are diverse and ingenious, and are only beginning to be understood.

Second, application of the knowledge obtained in these basic investigations promises to have a major impact on world health. Of the six diseases singled out for research emphasis by the Special Program for Research and Training in Tropical Diseases, sponsored by the World Bank, the United Nations Development Program, and the World Health Organization, five are caused by parasites: malaria, schistosomiasis, filariasis, trypanosomiasis, and leishmaniasis. These and other parasites affect more than a billion people worldwide. Approximately 300 million suffer from malaria, 200 million have schistosomiasis, and 300 million have filariasis. Another 40 million have onchocerciasis, which causes "river blindness."

Parasite diseases also affect the United States. Giardia, a protozoan parasite, is one of the most common causes of epidemic infectious diarrhea in this country. In addition, we must consider the immigration of a large number of people with parasitic diseases, the vast outflow of American tourists and business representatives overseas,

Copyright © National Academy of Sciences. All rights reserved.

and the potential exposure of military personnel in countries where parasitic infections are common.

The potential rewards of studying parasites are enormous, both in terms of learning about basic biologic processes and in combating disease. Because of their complexity, animal parasites are difficult to study. However, modern biologic concepts and advances in techniques are now enabling researchers to address such exciting and long-standing questions about parasite-host interactions as the following: (1) How are these unicellular and multicellular organisms capable of establishing infections? (2) What genes regulate the transformation of parasites through their complex life cycle? (3) What are the mechanisms of disease syndromes that result from parasitic infections? (4) What is the array of possible host protective responses and can these be enhanced? (5) What mechanisms do parasites use to evade the host's immune and other protective defenses? (6) Are vaccines feasible? (7) What biochemical pathways do parasites have that differ from the host and can these be used as targets for new drugs? (8) What is the basis for the ability of certain insects to transmit parasites or to become resistant to insecticides, and can these properties be altered?

Some of the technical advances that have been crucial to the study of parasites are the production of monoclonal antibodies, the isolation of specific genes, and the culturing and growth of some of these organisms in the laboratory. This paper describes some basic biologic questions that are being studied in parasites by several different scientific disciplines, and then suggests potential future applications of such information to prevention and treatment of parasitic diseases.

BASIC RESEARCH ON PARASITES

Molecular Biology

Antigenic Variation

African trypanosomiasis is characterized by cyclic parasitemia, that is, by the appearance in the blood stream every 7 to 10 days of waves of organisms. The surface of the protozoan's membrane is covered by a single type of glycoprotein, which is antigenic, and the parasites of each cycle contain a completely different glycoprotein. Although the body mounts an antibody response, the parasite can change its surface antigen hundreds of times, always keeping ahead of the host's immune response.

What is the basis for the capacity of these parasites to alter their surface antigens? Investigation of this phenomenon, undoubtedly the most sophisticated mechanism yet devised for evading the immune response of the host, uncovered the first example of "jumping genes" for surface proteins. Each organism has an estimated 300 to 1,000 different genes coding for these variant surface glycoproteins, but only one is expressed at a time. The gene to be expressed is duplicated and then moved to another part of the chromosome where it is expressed. It is not yet known how these parasite genes are selected, how they are turned on and off, or how they are regulated. Nevertheless, studies of the molecular biology of the African trypanosome already have had considerable influence on our understanding of gene rearrangement in general, and promise to uncover mechanisms of gene regulation.

Surface Antigen Repeating Units

Sporozoites, the form of the malaria parasite that matures in the salivary gland of the anopheles mosquito and is injected into the human skin during a bite, are free in the human host's blood for only a few minutes before entering liver cells. Because of their precise specificity for a single antigen, monoclonal antibodies have allowed the detection and characterization of surface proteins on sporozoites. The antibodies are specific to a given species as well as to that stage in the life cycle. The antibodies directed against these surface antigens neutralize sporozoite infectivity, so there has been great interest in both the antibodies and the antigens they have identified.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Genes coding for the sporozoite surface antigens have been introduced into bacteria by recombinant DNA techniques. This allows the production of sufficient antigen to study it in detail. The antigens have been found to have unusual repeating sequences. In the sporozoite of the malaria species *Plasmodium knowlesi*, for example, the unit is made up of a peptide with 12 amino acids repeated 12 times in tandem. The antigen of *Plasmodium falciparum* has a different repeating unit, with 4 amino acids repeated 23 times. It will be interesting to learn how common such repeating antigen units are in other parasites and what evolutionary pressures may have led to their development. In a more practical vein, such small reactive units lend themselves to the production of synthetic peptide antigens that may be used to develop vaccines effective against malaria.

Kinetoplast DNA

Kinetoplast DNA (kDNA) is a form of mitochondrial DNA found in some protozoans such as trypanosomes and leishmania, which is essential for their survival. It has a remarkable structure consisting of thousands of DNA circles interlocked in a network. A minor portion of these circles, the maxicircles, code for mitochondrial proteins. The majority, the minicircles, are only about 1,000 base pairs in size. These do not appear to code for any protein, and their function is not known. Minicircles are the only DNA in nature known to have a region of bent DNA helix. It is of great interest to learn both the function of this unusual DNA conformation and how its nucleotide sequence induces the curvature.

Gene Regulation During Life Cycle Transformations

Parasites undergo profound changes during the various stages of their life cycles, and work on the molecular biologic basis of these changes is in its infancy. The next few years should see dramatic progress in work on the molecular basis for developmental transformation in many different protozoan and helminthic parasites.

Leishmania, for example, go from a flagellated protozoan form in the insect host to a smaller form without flagella when in a mammalian macrophage. This change is reversed after the protozoa are taken up by a fly when it bites an infected animal. The amount of tubulin, a structural protein of the cytoskeleton, decreases as the parasite goes from the flagellated to the nonflagellated form and then increases when the parasite returns to the flagellated form. In this case of transformation, the control of the genes for tubulin occurs at the level of messenger RNA processing.

Immunology

Studies in immunology have concentrated in several areas. These include mechanisms the host can muster that the parasite cannot escape, mechanisms of immune evasion, and the role of the immune response in causing tissue damage. The information obtained has been useful in developing better diagnostic reagents and forms the basis for the development of protective vaccines.

Protective Mechanisms of the Host

Studies on host responses to parasites have revealed some novel systems. For example, one such system was found while studying the antibody-dependent ability of host cells to kill schistosomes. These studies led to the first demonstration that the eosinophil, a white blood cell, can act as a killer cell. These findings have now been applied to other areas where it has been shown that eosinophils play an important part in producing tissue damage in certain types of heart disease, in inflammation, and in allergic diseases such as in asthma.

As another example, work on the immune response to ticks has highlighted the role of basophils, another type of white blood cell, and mast cells in producing vascular perme-

Copyright © National Academy of Sciences. All rights reserved.

ability in the early stages of cell-mediated immune reactions. Furthermore, studies of the schistosome-linked granuloma (an inflammatory nodule) were the first to demonstrate suppression of lymphocyte hormone production by T-cells. (T-cells and lymphocyte hormones are important regulatory components of the immune system.)

Mechanisms of Immune Evasion

Parasites have developed many ways of evading the human host's defenses. Some change their surface antigens, some masquerade as the host by taking up host molecules onto their surfaces, and others simply shed their surface antigens. Some have enzymes on their surface that destroy antibodies. Macrophages are the body's major cell for engulfing foreign particles, and they usually kill the organisms they engulf. Some parasites have devised ways of avoiding the effects of the toxic substances present in the lysosomes of the macrophage, although how they do this is not known. Yet other parasites suppress the host's immune attack.

Although at present most of our knowledge is simply descriptive, future studies should elucidate the mechanisms for accomplishing all these forms of evasion. Information about the effects of parasites on the immune system would undoubtedly enhance our understanding of how the immune system functions in many other types of diseases.

Immunopathology

Studies on the mechanisms of tissue damage in schistosomiasis demonstrated that the host reactions to egg antigens are the main factors producing pathology. Further studies on this process were the first to show that mononuclear cells of the host release factors that promote fibrosis. The mechanisms of pathology induced by many different parasites merit further study. Of particular interest are the ways parasites induce autoimmune reactions such as those postulated in infections caused by *Trypanosoma cruzi*, the mechanism of blindness caused by *Onchocerca volvulus*, the process of sequestration of infected red blood cells leading to malaria damage in the brain, and the underlying basis for destruction of cartilage in mucocutaneous leishmaniasis.

MEMBRANE BIOLOGY AND CELL BIOLOGY

Unusual Features of Parasitic Membranes

The first interactions between the parasite and the host are membrane-membrane interactions. Many parasites have membranes that are different from those in mammalian cells, and considerable effort has been directed at learning about the function and structure of parasite membranes. As a result, several interesting discoveries have emerged.

One example is the finding that the variant surface glycoprotein of African trypanosomes is attached to the membrane by a novel protein-lipid bond. A newly recognized enzyme is involved in destroying this bond. Another unusual finding concerns an enzyme on the surface membrane of *Trypanosoma cruzi*. This organism has a neuraminidase, an enzyme that cleaves off the most important sugar group on the surface of mammalian cells, namely, sialic acid. The role of this enzyme should be studied, especially as it relates to parasite survival and tissue injury. The question of whether similar enzymes are present on other parasites also is of interest.

Parasite-Host Membrane Interactions

Work is being carried out on the mechanism by which malarial merozoites (the mobile infective stage) invade red blood cells. Glycophorin, a glycoprotein on the red blood cell surface, appears to be an important part of the receptor for the parasite. How parasites disrupt the rigid cytoskeleton of the red blood cell so they can enter it is unknown, as is the

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

source for the special membrane that surrounds the parasite once it is in the red blood cell. Furthermore, parasite antigens soon appear on the outer surface of the red blood cell. To reach the surface, parasite antigens must traverse the parasite's own membrane, then the special membrane surrounding the parasite, and finally the red blood cell membrane. The mechanism for accomplishing this voyage is unknown.

Alteration of the red blood cell surface is important in enabling the red blood cells that contain the parasite of *Plasmodium falciparum*, the most common of the four malaria species, to lodge in the vascular bed of certain organs such as the brain. These surface antigens also undergo antigenic variation by unknown mechanisms. The spleen can alter the expression of these parasite antigens on the red blood cell surface, but how it does so is another mystery to be solved.

Of interest is the recent discovery that *Entameoba histolytica*, another protozoan parasite, kills host cells by injecting a protein into the host cell membrane that causes an ion flux and subsequent osmotic lysis of the target cell. These findings have stimulated a search to see whether analogous proteins are involved in cytotoxicity induced by T-lymphocytes.

Clearly, studies on these membrane effects will have ramifications affecting much more than our knowledge of the parasites themselves.

Biology of Insect Vectors

An attack on the insects that transmit particular pathogens remains a major focus for most programs aiming to control parasitic diseases. Although chemical insecticides provide the major tool for this purpose, their effectiveness has steadily diminished because of insect resistance.

New technologies should lead to revised strategies for combating insect vectors. Recent developments in molecular biology, combined with improved understanding of insect physiology and behavior, should enable us to better understand the mechanisms of insecticide resistance, vector competence—the ability of an insect to transmit a parasite—and the control, feeding behavior, and reproduction of insects. Ultimately, this technology may enable us to genetically modify insect populations in nature, so that the targeted population is less capable of transmitting pathogens or less susceptible to becoming resistant to insecticides.

Such research, both basic and applied, should have far-reaching benefits, affecting not only the transmission of known disease but also important aspects of agriculture.

Biochemistry and Pharmacology

The very nature of parasitism implies that the parasite has metabolic needs that the host supplies and, therefore, that the parasite must have metabolic pathways that differ from those of the host. Novel metabolic pathways and even new cell organelles are being discovered. These have led to the design of compounds that can inhibit these unique pathways and thus control some parasites and insects without affecting their hosts. A few examples are described below.

Unusual Metabolic Pathways

In general, parasitic protozoans cannot synthesize purines, which are chemical precursors of nucleic acids. They have therefore developed elaborate pathways to utilize purines from the host. Many of these pathways differ from one protozoan to another. Study of these pathways has greatly enhanced our knowledge of purine metabolism in general and may lead to development of drugs that can kill protozoans but are not toxic to mammalian cells. Polyamines also are required for DNA replication and cell differentiation. Trypanosomes and leishmania have

Copyright © National Academy of Sciences. All rights reserved.

simplified metabolic pathways for biosynthesis of polyamines, which may be susceptible to specific inhibition.

Unusual Organelles

Parasites have been found to contain several novel organelles. One of these, the glycosome, found in trypanosomes, contains the enzymes for glycolysis; its discovery encourages research toward development of specific inhibitors. Another novel organelle is the hydrogenosome found in trichomonas. This organelle contains pyruvate kinase and pyruvate dehydrogenase, two enzymes used in producing the ATP needed by the cell. These hydrogenosomes chemically reduce the drug metronidazol to a toxic compound that kills the parasite. The host cells, which lack this organelle, are spared.

The Neuromuscular Junction

The neuromuscular junction of certain helminths contains receptors for GABA (gamma-aminobutyric acid), which also is a neurotransmitter in the human brain. Ivermectin, a drug that enhances the activity of these receptors, has a profound effect on helminths. It causes paralysis that prevents them from moving or eating. It is exceedingly potent, with less than milligrams per acre of pastureland protecting cattle against helminth parasites. Because this drug cannot traverse the blood-brain barrier, it is nontoxic to humans. Further studies using ivermectin have shown that it also acts on the GABA receptor complex of arthropods. Additional studies with this drug should greatly increase our understanding of insect physiology and yield knowledge of GABA receptors in higher organisms.

Juvenile Hormone

Juvenile hormone (ecdysin) is involved in the differentiation and molting of insects and has recently been found in helminths. Further study of the regulation of molting and differentiation of helminths and the role of ecdysin in this process should lead to entirely new approaches to control of some of these parasites.

APPLICATION TO DISEASES

Increased understanding of the basic biology of parasitism should provide many opportunities for combating diseases caused by these organisms. This is especially important now that traditional public health measures that had helped to control some of these diseases are no longer sufficient by themselves. An example of the need for new methods can be seen in the case of malaria. After World War II, there was great hope that this disease could be eradicated by using DDT to kill the mosquitoes in houses and chloroquin to treat infections in people. Indeed, malaria was eradicated from the United States and around the Mediterranean and decreased markedly in Asia and South America. In the early 1960s in Sri Lanka, the number of cases per year fell from more than 1 million to 18. However, within 5 years, there were again almost a million cases, due to a combination of DDT and chloroquin resistance and to the difficulty in maintaining this costly control program. Now malaria has reached serious epidemic proportions in parts of Asia and South America and continues unabated in Africa where it was never controlled due to the behavior of the local mosquitoes, which, instead of landing on the sprayed walls of the house after biting, go outside. Malaria resistant to the present drugs has become an alarming problem.

The belief that parasitic diseases disappear with modernization and industrialization is not always correct. For instance, the prevalance of schistosomiasis has increased hand in hand with the development of hydroelectric dams required for energy and irrigation projects necessary for improved agriculture. The large lakes behind these dams have added thousands of miles of waterfront

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

and have increased the contact between people and the infected snails that transmit this parasite. Another example is the increase in leishmaniasis in the Amazon where people who are expanding towns and building roads come into contact with sandflies that live in the forest and transmit this infection. Applying insecticides to the forests is neither effective nor feasible.

FUTURE RESEARCH OPPORTUNITIES

There is a great need now to broaden the entire base of biologic research on parasites. This is an opportune time because many of the new concepts and biotechnologies make it possible to solve some of the important questions concerning the biology of parasitism. A multidisciplinary approach is required.

Most of the productive research thus far employing the new biotechnologies has been on two parasites, the malaria parasite *P. falciparum* and the African trypanosome of cattle, *T. brucei*. Much less work in molecular biology, for instance, is being done on the African trypanosomes that infect humans, or on other protozoans and the many helminths.

The main reason for the early focus on *P. falciparum* and *T. brucei* is the relative ease, compared with other parasites, with which these two parasites can be cultured and manipulated in the laboratory. There is much excitement over prospects of a vaccine against the infective form of *P. falciparum*, the sporozoite, but little work has been carried out on the other three species of malaria that infect humans. There is also work in progress on the development of vaccines to the other stages of the *P. falciparum* parasite because it is still not clear how effective the vaccine against the sporozoite will be. The sporozoite vaccine has worked in laboratory tests on rodents, but must now be developed for humans. One problem is that the vaccine will have to work completely in the short time that the sporozoite is present in the blood. An effective adjuvant must be identified, one that will allow for a strong enough response for the vaccine to be clinically effective.

It is now important to take advantage of new technologies to study other parasites. This will require establishing cultures of these organisms in vitro, developing suitable animal models, and establishing appropriate life cycles in the laboratory. Such work is time-consuming, may involve many false starts, and requires cooperation of persons in many disciplines and long-term support.

Molecular biology is providing several new methods for generating antigens needed to form the basis of vaccines. These methods include laboratory synthesis of peptides, producing proteins in microorganisms by recombinant DNA techniques, and incorporation of genes coding for protective parasite antigens into the DNA of attenuated viruses used for unrelated vaccines. The essence of the problem is no longer how to produce antigens by recombinant techniques, but which antigens to produce and how to present them effectively to the immune system in clinically acceptable formulations.

Molecular biologic techniques also are being used for better diagnosis. For instance, the nucleotide sequence of the kinetoplast DNA found in certain protozoans varies among the different species. Using DNA hybridization to identify the kDNA, it is now possible to determine in 24 hours instead of 3 months whether a leishmania infection involves a benign or virulent species. Similar techniques should be applicable to other parasites. Furthermore, the production of species-specific monoclonal antibodies against parasites causing malaria, leishmaniasis, schistosomiasis, and other diseases is leading to the development of greatly improved diagnostic reagents and identification of relevant antigens for the development of vaccines.

The study of the genes that regulate the transformation of the parasite through its different stages should result in novel ways of interrupting the life cycle. Studies of the molecular biology of parasites have already broadened our understanding of molecular

Copyright © National Academy of Sciences. All rights reserved.

biology in general. For instance, it was through the study of malaria that an enzyme, mung bean nuclease, was found recently to have the unique property of cutting genes out of DNA without cutting in the coding region, so that genes can be studied and cloned much more easily than before.

Basic immunologic studies of parasites should be expanded to cover several other areas beyond development of vaccines and diagnostic reagents. For instance, the study of the mechanisms of immune evasion should lead to new methods of overcoming the parasites, and those methods should also be adaptable to such other types of organisms as bacteria, viruses, and fungi. Work on immunopathology, especially possible autoimmune mechanisms, is required to assure that the antigens proposed for vaccines are not also those that can induce pathology detrimental to the host. This type of study is important, for instance, in *T. cruzi* infections, because it has been postulated that the heart disease resulting from this infection may be caused by an autoimmune process.

Basic research on the novel biochemical mechanisms that parasites have evolved is currently one of the most neglected aspects of parasitology. Work in this area should lead to the development of new drugs that are needed to treat some of these diseases. For instance, the major drug for treating African trypanosomiasis is Suramin, first synthesized in 1917. It is a very toxic drug and is effective only in the early stages of the disease, before the parasite reaches the brain and produces the symptoms that give the disease its common name, sleeping sickness. There are no drugs to treat *T. cruzi* infection, and those available against the leishmania organisms were first produced decades ago and are quite toxic.

Studies of the biology of insects that carry and transmit parasites should be expanded to include the application of molecular biology. Alternative methods of insect control, including the genetic modification of insect populations in nature, should result. Investigations in this area should not only help control some of these parasite diseases but also should be applicable to problems in agriculture and animal husbandry.

Despite the enormous potential of research on the basic biology of parasites, relatively few scientists are working in this area, and many more should be encouraged to do so. The foregoing examples of present work on the biology of parasitism should make clear the excitement of imminent discovery and the promise that the research not only will reap new scientific knowledge, but also will enable applications of immense practical consequences. Further study offers tremendous opportunity for progress.

Copyright © National Academy of Sciences. All rights reserved.

Report of the Research Briefing Panel on Solar-Terrestrial Plasma Physics

Copyright © National Academy of Sciences. All rights reserved.

Research Briefing Panel on Solar-Terrestrial Plasma Physics

Charles Kennel (*Chairman*), University of California at Los Angeles, Los Angeles, Calif.
Jonathan Arons, University of California at Berkeley, Berkeley, Calif.
Roger Blandford, California Institute of Technology, Pasadena, Calif.
Ferdinand Coroniti, University of California at Los Angeles, Los Angeles, Calif.
Martin Israel, Washington University, St. Louis, Mo.
Louis Lanzerotti, AT&T Bell Laboratories, Murray Hill, N.J.
Dennis Papadopoulos, University of Maryland, College Park, Md.
Robert Rosner, Harvard University, Cambridge, Mass.
Frederick Scarf, TRW Systems, Redondo Beach, Calif.

Staff

Donald C. Shapero, *Staff Director*, Board on Physics and Astronomy, National Research Council
Charles K. Reed, *Rapporteur*, Board on Physics and Astronomy, National Research Council
Helene Patterson, *Staff Assistant*, Board on Physics and Astronomy, National Research Council
Allan R. Hoffman, *Executive Director*, Committee on Science, Engineering, and Public Policy

Copyright © National Academy of Sciences. All rights reserved.

Report of the Research Briefing Panel on Solar-Terrestrial Plasma Physics

1 INTRODUCTION

Radio waves were first reflected from the earth's ionosphere—the very edge of space—in the 1920s. Plasma waves were discovered in the laboratory at about the same time. Soon thereafter scientists began to develop a new branch of physics—plasma physics*—to answer the questions stimulated by observations such as those above. Why do laboratory gas discharges exhibit unexpected collective behavior? How do radio waves propagate in the ionosphere? Why does the appearance of sunspots on the sun presage magnetic storms and auroral displays at earth? What roles do magnetic fields play in stars and galaxies? By the late 1940s, physicists had concluded that plasma is a distinct, fourth state of matter, with properties that are unique because cooperative long-range interactions among the charged particles are far more important than collisions. Moreover, it had become clear that, except for occasional cool regions like planets and their atmospheres, most of the universe is in the plasma state.

Modern plasma physics began in the 1950s. The initiation of research aimed at harnessing the energy source of the stars—thermonuclear fusion—and the first launch of an artificial earth satellite catalyzed the development of the technology needed to study hot plasmas in the laboratory and natural plasmas in space. Moreover, the remarkable discoveries of the solar wind and of the Van Allen radiation belts demonstrated that plasma physics is essential to understand the space environment of the

This research briefing is based on the report of the Subpanel on Space and Astrophysical Plasma Physics of the *Survey of Physics* (in preparation [see App. B]), and on the report of the Committee on Solar and Space Physics of the Space Science Board entitled *Solar System Space Physics in the 1980's; a Research Strategy* (1980). This briefing was reviewed and endorsed by the Committee on Solar and Space Physics on August 1, 1984.

*Plasma physics combines concepts from electromagnetism, fluid physics, statistical mechanics, and atomic physics into a unified methodology for the study and practical use of the nonlinear collective interactions of charged particles with each other and with electric and magnetic fields. All physical conditions involving thermal energy densities exceeding 1 electron volt per atom must use plasma physics. A fluid description, magnetohydrodynamics, is used to describe the large-scale properties of plasma systems, and plasma kinetic theory is used to describe the microscopic processes that influence the large-scale properties.

Copyright © National Academy of Sciences. All rights reserved.

earth and its effects on civilian and defense systems operating in this environment.

The two principal applications of contemporary plasma physics are to fusion research and to solar-terrestrial research—the study of the chain of physical processes that starts with the generation of the sun's magnetic field in the solar interior and links it to activity at the sun's visible surface and, ultimately, to the earth's ionosphere and atmosphere (Section 2). Our understanding of the plasma processes occurring in this interaction chain has made rapid progress in the past decade. The increasing precision of measurements, numerical modelling, and theory applied to solar-terrestrial plasma problems amounts to a revolution in technique relative to 10 years ago. The experimental diagnosis and theoretical interpretations of many space plasma processes now match in precision the best of current laboratory practice. As a result, the study of space plasmas has also become one of the most active areas of basic plasma research.

Research on solar-terrestrial plasmas is advancing our theoretical understanding of many basic plasma processes—for example, how convective turbulence generates magnetic fields, how plasma systems adjust their structure and dynamics by the process of magnetic reconnection, and how particles are accelerated to high energies. These and many other space plasma processes are nonlinear, and their study is contributing to a new branch of mathematical physics—nonlinear dynamics—which has recently led to exciting advances in many branches of physical science.

An increasingly clear perception of the functions of the links in the solar-terrestrial interaction chain has led to a design for a unified study of the solar-terrestrial system as a whole—the proposed multispacecraft International Solar-Terrestrial Physics Program (Section 3). Together with the Solar Optical Telescope, discussed in Section 2, this program would be the fundamental underpinning of solar-terrestrial plasma research for the next 10 years. We recommend that the United States commit to participating in the International Solar-Terrestrial Physics Program, by initiating funding in FY 1986.

This research briefing panel has the special responsibility to point out new opportunities. In Section 4, we will discuss the growing impact of solar-terrestrial research on astrophysical plasma physics. The plasma phenomena in the solar system have proven to be examples of processes that also occur in other stars, near neutron stars and black holes, and in galaxies. The sun and solar system have become a laboratory in which astrophysical plasma processes can be studied *in situ* and with a precision attainable nowhere else. As a result, space and astrophysical plasma physicists have begun to work closely together over the past 5 years, and a new and broad research field is beginning to develop.

The great importance of numerical computations to all plasma physics is taken up in Section 5. Such computations have already critically advanced research on fusion plasmas, and we are confident they can be equally important to solar-system and astrophysical plasma physics. We recommend early initiation of a program of large-scale numerical computations dedicated to these subjects, because detailed mathematical models clarify fundamental understanding, strengthen the interpretation of observations, and, in the case of the solar-terrestrial interaction chain, might ultimately lead to some predictive capability. Such a program would also encourage organizationally the already fruitful interactions between space and astrophysical plasma research.

2 THE SOLAR-TERRESTRIAL SYSTEM

Research in this century has revealed a chain of interactions, almost all of which involve plasma physics, that connects the

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

generation of the sun's magnetic field, activity at the surface of the sun, the generation of the solar wind, and the dynamic behavior of the earth's magnetosphere and upper atmosphere.

The first link in the interaction chain is the sun's magnetic field. Historically, the subject of solar-terrestrial physics began with the observation that auroral and magnetic activity at earth is correlated with the 22-year cycle of the solar magnetic field. Today, no problem in solar-terrestrial research presents a greater challenge than understanding the generation and dynamics of the solar magnetic field. How does convective turbulence below the visible surface of the sun create the sun's weak magnetic field and the persistent concentrations of magnetic field that are sunspots? Recent observations of the microstructure of the sun's magnetic field challenge our understanding of solar turbulence: almost all the surface field is concentrated into tiny regions covering only 1/1000 of the surface area. The ultimate goal remains to understand how the remarkably well-ordered solar cycle arises out of such turbulent chaos.

New observational tools are at hand to help answer some of the challenging questions before us. By observing the periodic motions at the surface of the sun, the new technique of solar seismology is now revealing for the first time the structure of the solar convection zone, at the depths within the sun where the solar magnetic field is created. The Solar Optical Telescope, scheduled for a Shuttle flight in 1990, will measure photospheric and chromospheric structures with the extraordinary spatial resolution of 70 km, providing hitherto unobtainable information essential to theories of magnetohydrodynamic turbulence and solar-surface magnetic fields and dynamics. Later, this instrument is scheduled to become the principal component of the Advanced Solar Observatory, to be flown on a continuing basis on an orbiting platform.

The turbulence in the sun's outer layer heats the tenuous plasma in the solar corona to a temperature of several million degrees; however, the exact mechanism by which the plasma is heated remains uncertain. Ten years ago, *Skylab* observations revealed that the magnetic field emerging from the visible surface evolves into a complex set of open and closed coronal structures that persist for several solar rotation periods (27 days) and mysteriously resist destruction by the sun's differential rotation. The closed field regions—coronal loops—appear to channel energy into the corona and keep it there until microscopic plasma processes dissipate it. These regions are the sites of many small flares, and occasionally of a giant flare, in which magnetic energy is converted into heat and energetic particles, perhaps by a magnetic reconnection process. On the other hand, the energy flowing outward on open field lines drives a plasma wind that expands into interplanetary space. This solar wind carries hot coronal plasma and the sun's open magnetic flux past all the planets of the solar system, interacting with each to form a distinctive magnetosphere.

Our next challenge is to increase our understanding of coronal structure and plasma dynamics and of the processes regulating the flow of energy at each level of the sun's atmosphere. We hope ultimately to integrate this understanding into a unified model that accounts for coronal heating, solar flares, and the generation of the solar wind.

The solar wind is a supersonic, strongly ionized flow that carries plasma, energy, angular momentum, and magnetic field throughout the solar system, interacts with the planets and comets, and fills a three-dimensional volume of space called the heliosphere. The heliosphere terminates where the solar wind is decelerated to subsonic speeds by its interactions with interstellar matter. The solar wind is an excellent laboratory for the investigation of plasma processes of importance to basic

Copyright © National Academy of Sciences. All rights reserved.

plasma physics, solar-terrestrial research, and astrophysics. For example, fully developed magnetohydrodynamic turbulence can be conveniently studied in the solar wind but not in the laboratory. The variability of the solar corona creates a variety of structures propagating in the solar wind whose study has enriched our understanding of magnetohydrodynamics. One of these, the collisionless interplanetary shock, which triggers magnetic storms at earth, accelerates energetic particles by processes thought similar to those that accelerate galactic cosmic rays. The cosmic rays from the galaxy diffuse into the solar system by scattering from solar wind turbulence. As a result, the galactic cosmic-ray flux received at earth is largest at the minimum activity phase of the solar cycle.

Our next challenge is to synthesize our growing understanding into quantitative models that relate coronal activity to the three-dimensional behavior of the solar wind and its propagating structures. To do so, we will need to understand the microscopic processes that regulate plasma transport and the acceleration and propagation of energetic particles.

The solar wind is the plasma link between sun and earth. The solar wind does not directly hit the earth, but is largely deflected by the earth's magnetic field; the name magnetosphere refers to the region enclosing the earth in which the geomagnetic field organizes the behavior of the plasma. Within the magnetosphere, the geomagnetic field traps energetic particles to form radiation belts whose intensity is regulated by microscopic plasma turbulence. Solar-wind energy is coupled into the magnetosphere by the process of magnetic reconnection. When it occurs in the magnetopause, the boundary layer separating the solar wind from the magnetosphere, the reconnection opens part of the earth's magnetic field to the solar wind and stretches the open field into a long magnetic tail extending hundreds of earth radii downstream. A compensating reconnection process in the magnetic tail converts magnetic energy into particle energy and generates strong magnetohydrodynamic flows, which are directed away from the earth tailward of the reconnection region and towards the earth's night side earthward of the reconnection region. Although the consequences of reconnection are clear, the plasma physics of reconnection is not well understood. In the magnetopause, reconnection appears to be spatially structured and temporally variable. In the geomagnetic tail reconnection can be either steady or explosive. One thing is certain: the variability of the solar wind stimulates violent, unsteady magnetospheric flows, which lead to events called substorms, whose effects are felt in the auroral ionosphere and throughout the magnetosphere and tail.

The earth's magnetic field couples the flow in the magnetosphere and tail to the ionosphere and upper atmosphere. Nonlinear plasma processes associated with the electrical currents directed along the field lines connecting flowing magnetospheric plasma to the ionosphere accelerate electrons to about 10 keV. These electrons create the aurora when they hit the atmosphere. The same processes also heat ionospheric plasma, which then boils back out into space to modify the magnetohydrodynamic flow. Some scientists have suggested that the solar-terrestrial interaction chain reaches deep into the atmosphere to modify terrestrial weather and climate; however, the mechanisms by which this might occur have not been identified.

Thus far, solar-terrestrial research programs have concentrated on elucidating individual links in the causal chain connecting the solar wind to the magnetosphere and ionosphere. Our challenge today is to increase our understanding of each of the local plasma processes influencing the magnetosphere's structure and dynamics so that we can begin to create quantitative models that start with the solar wind's

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

observed state and calculate the magnetosphere's and ionosphere's responses.

The most spectacular manifestation of the solar-terrestrial interaction chain is the magnetic storm. The first evidence that a large solar flare might occur is the appearance of a complex sunspot group in the sun's photosphere. Prompt electromagnetic radiation arrives at earth a few minutes after the energy in the coronal magnetic fields associated with the sunspot group is suddenly released. Energetic solar-flare protons are guided by the solar wind and magnetospheric magnetic field into the earth's polar atmosphere soon afterward. The enhanced ionospheric plasma produced by the energetic protons attenuates the radio noise received from cosmic radio sources. Several days later, a shock wave passes over earth, enveloping it in dense, hot solar-flare plasma that compresses the magnetosphere. Substorms increase in frequency and strength, expel hot plasma tailward, and inject hot plasma into the earth's equatorial inner magnetosphere to form a "ring current," which creates the geomagnetic field depression and activity that first motivated the name magnetic storm. The aurorae intensify and move to low latitudes, creating a dense highly disturbed ionospheric plasma that interferes with radio communications. Intense wind systems, sometimes of a worldwide scale, are generated in the upper atmosphere.

Many practical systems, both civilian and defense, and all our manned space endeavors must operate in the highly variable and potentially hostile plasma environment of the earth and solar system. Plasma processes in this environment also influence and even disrupt important ground-based systems over local and regional scales. Ground-based high-frequency communication systems in the earth's polar regions can be entirely blacked out by magnetic storms. Spacecraft have been electronically disabled by violent electrical discharges that occur when hot plasma from the magnetic tail envelops the spacecraft. The risk of such disasters can be reduced only by continuing attention to the effects of the plasma environment on spacecraft systems. It is clear that to work in the space environment, we must understand it.

3 THE INTERNATIONAL SOLAR-TERRESTRIAL PHYSICS PROGRAM

There is general agreement within the research community about the objectives of Solar-Terrestrial Research (Appendix A). Furthermore, there is remarkable unanimity that we can now begin the quantitative study of the solar-terrestrial system as a whole. The unanimity is expressed in the proposal currently before the administration to initiate U.S. participation in the International Solar-Terrestrial Physics Program (ISTP). Spacecraft provided by the National Aeronautics and Space Administration (NASA), the European Space Agency (ESA), and the Japanese Institute for Space and Aeronautical Sciences (ISAS) would make simultaneous, coordinated, detailed measurements of key links in the solar-terrestrial chain. The ISTP program would consist of six elements. Two spacecraft would study the sun and solar wind—ESA's Solar and Heliospheric Observatory (SOHO) and NASA's Wind spacecraft. Three would study critical regions of the earth's magnetosphere identified by previous research—POLAR (NASA), EQUATOR (NASA), and GEOTAIL (ISAS) (their names indicating the regions to be studied). ESA's multiple spacecraft CLUSTER would study small-scale plasma turbulence and boundary layers in the magnetosphere. Modern communications and data processing technology will make it possible for scientific investigators to obtain a unified set of data from all ISTP spacecraft, a prerequisite to quantitative understanding of the solar-terrestrial system as a whole.

The ISTP draws on the experience of previous coordinated programs, such as the

Copyright © National Academy of Sciences. All rights reserved.

International Geophysical Year (1957-58) and the International Magnetospheric Study (1977-80).* The ISTP is a central part of a more comprehensive effort that includes other funded spacecraft, such as the International Solar-Polar Mission (NASA-ESA), the Upper Atmospheric Research Satellite (NASA), and EXOS-D (Japan). In addition, many ground-based facilities, airborne observatories, balloon and rocket campaigns, and other spacecraft can provide data for collaborative research projects. Active plasma experiments, carried out by Shuttle and other means, will enhance, and be enhanced by, the ISTP.

The United States has already provided much of the leadership in conceiving and designing ISTP, and ISTP will enable the United States to secure a position of future leadership. If it is successful, the ISTP data network, and the collaborative mode of research engendered by its existence, might well outlive the original ISTP spacecraft and shape the way that future solar-terrestrial research will be carried out.

Only a coordinated program, such as ISTP, can study the complex interaction between the parts and the whole of the solar-terrestrial system—an interaction characteristic of all large-scale plasma systems. By studying individual links in the interaction chain, ISTP will provide qualitatively new measurements critical to understanding many fundamental plasma processes—such as dynamo magnetic field generation, particle acceleration, and magnetic field reconnection—that are also important to both laboratory and astrophysical plasma research. By studying several links in the chain simultaneously, ISTP will define the contexts in which these processes occur, a prerequisite for quantitative understanding. By studying the system as a whole, ISTP will motivate and test a series of increasingly comprehensive models, whose goal is first to understand, and then to predict, the highly variable plasma environment of the sun, solar wind, and earth. More and more practical systems will have to function in this plasma environment. For all the above reasons,

We recommend: that the United States initiate its participation in the International Solar Terrestrial Physics Program as scheduled in FY 1986.

4 THE IMPACT OF SOLAR-TERRESTRIAL PLASMA PHYSICS ON ASTROPHYSICS

Recent events have reaffirmed the essential unity of laboratory, solar-system, and astrophysical plasma physics that was perceived by the pioneers of the subject 50 years ago. The three branches of plasma physics are unified by a shared preoccupation with a common set of basic physical problems (Appendix B). Yet, during the past 25 years, each branch developed more or less independently, as each individually explored the opportunities of technique and application inspired by modern laboratory, space, and computational technology. A new trend is beginning to emerge. The realization is growing that solar-terrestrial research has developed problems in plasma physics that not only are fundamental to basic plasma physics but are broadly applicable to many astrophysical systems. In particular, a new field of space and astrophysical plasma physics is developing that takes as its starting point the great advances in all three branches of plasma physics during the past 25 years.

The first successes of this new, broader discipline are already in evidence. Terrestrial magnetospheric physics provided a coherent framework for understanding the

*Neither the solar-terrestrial interaction chain nor the role of plasma processes in it were well defined at the time of the International Geophysical Year, and the International Magnetospheric Study did not include the sun and solar wind, and so did not attempt to study the whole interaction chain.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

magnetospheres of Mercury, Jupiter, and Saturn. The upcoming spacecraft encounters with Uranus and Neptune will further test and extend magnetospheric concepts. Since Venus is unmagnetized, studies of Venus' interaction with the solar wind will provide a foundation for investigating the plasma physics of cometary magnetospheres, which soon will be studied by spacecraft for the first time.

Magnetic fields thread their way among planets, stars, galaxies, and probably even clusters and superclusters of galaxies. By transporting angular momentum, turbulent magnetic fields may regulate the accretion flow onto white dwarfs, neutron stars, and black holes. The gravitational energy released when accreting matter falls onto such compact objects may power the enormous energy outputs of galactic x-ray sources and quasars. The strong convective turbulence and buoyancy forces that develop in any gravitationally bound atmosphere will rapidly dissipate and expel magnetic fields. Hence, in nearly all astrophysical bodies, new magnetic fields must be continuously regenerated by turbulent magnetic dynamo processes. The sun provides the only accessible laboratory in which the dynamo mechanism in a highly conducting plasma can be observed and our theoretical dynamo models tested.

Plasma outflows like the solar wind are found throughout the cosmos. The *Einstein* spacecraft made the startling discovery that most stars have hot x-ray emitting coronae, implying that convective envelopes, magnetic dynamos, surface magnetic activity, and strong winds are common. Much of the interstellar medium is filled with hot, low-density plasma from blended stellar winds and supernova remnants. Since the densities, velocities, temperatures, and magnetic field strengths in the interstellar plasma are similar to those of the solar wind, solar system measurements are automatically relevant to the physics of the interstellar medium. Surrounding gas pressure collimates the winds from newly forming stars into bipolar jets; similar jets are observed in the exotic compact object, SS-433, and in radio galaxies. Relativistic winds, possibly consisting of electrons and positrons, are thought to flow away from pulsars and supermassive black holes in active galactic nuclei. The basic physics that underlies all these astrophysical winds was formulated and observationally confirmed in studies of the solar wind.

Super-energetic plasmas occur throughout the astrophysical universe. Quasars and active galactic nuclei convert prodigious amounts of energy into the relativistic plasmas that form the galactic scale radio jets. Trillion-volt electric potentials develop near the magnetic polar regions of pulsars. Radiation from the particles accelerated in these potentials is thought to trigger a pair-production cascade that fills the surrounding magnetosphere with an electron-positron plasma and powers a highly relativistic wind. In our galaxy, supernova shocks accelerate the cosmic rays up to a total energy density that is comparable with the thermal and magnetic energy densities of the interstellar medium. The plasma processes responsible for particle acceleration have long been studied in the solar-terrestrial context—in planetary magnetospheres, at planetary bow shocks and interplanetary shocks, and in solar flares. The acceleration models derived from these observations guide and test the underlying plasma physics of their astrophysical counterparts. For example, although it is not possible to detect the plasma processes responsible for particle acceleration in supernova shocks, theories of cosmic ray acceleration can be tested by measurements of solar system shocks together with measurements of galactic cosmic ray energy spectra and composition.

The challenge for the next decade is becoming clear: space and astrophysical plasma physicists must work side by side to formulate quantitative models of solar system and astrophysical phenomena in

Copyright © National Academy of Sciences. All rights reserved.

which magnetohydrodynamics and plasma physics are essential for the interpretation of observations, recognizing that most of the universe is in the plasma state.

5 LARGE-SCALE NUMERICAL MODELS

Plasma physics was a pioneer in the successful utilization of large-scale computations for fluid, magnetohydrodynamic, hybrid, and kinetic models. Most of the progress since the late 1960s has been in fusion research and nuclear weapons phenomenology. The establishment of a computational facility dedicated to magnetic fusion significantly advanced understanding of magnetic confinement systems and of fusion and basic plasma processes. The establishment, in 1979, of NASA's highly successful Solar-Terrestrial Theory Program made numerical models and simulations regular tools in solar system plasma research and has prepared that research community for the next, more advanced stage.

In the past decade, our research on solar-terrestrial plasmas has achieved a measure of quantitative understanding through the use of numerical models. The study of plasmas beyond the solar system has developed more slowly than space plasma physics for a fundamental reason: the microscopic plasma processes that regulate the behavior of astrophysical systems cannot be observed directly, as can those in the solar system and in the laboratory. Now, however, the modern theoretical and computational techniques developed to understand laboratory and space measurements have opened the door to modelling of the plasmas in the still larger and more exotic environments of astrophysics. Since the underlying astrophysical plasma processes cannot be directly detected, we believe that the best strategy will be to create numerical models at the large-scale system level that postulate small-scale plasma and radiation processes and to iterate between the system and process levels until quantitative agreement with observation is achieved. It is our perception that, for the first time, the current level of development of numerical technology, theory, and observations gives such a strategy a significant chance of success.

The continued development of numerical technology will advance many branches of science. We foresee that many problems of a scale that requires today's national computing facilities will soon be addressable by local university and laboratory facilities. This will only increase the importance of numerical modelling to our subjects. Nonetheless, we believe that the leading research on many solar system and astrophysical plasma models will continue to be done on the most advanced computing facilities existing at any given time, because these models involve a complex interplay between large- and small-scale processes.

Thus far, the responsibility for the maintenance and advancement of state-of-the-art computing facilities has been a national one, because it is beyond the capability of single institutions and because a national scope provides an adequate pool of users. America's existing advanced computational facilities, devoted to defense, fusion research, and meteorology, have been used on a piecemeal basis for solar system and astrophysical plasma research. These busy facilities do not have space physics and astrophysics as an institutional objective, and researchers in these fields must make individual agreements to secure access to advanced computing. In some cases, American researchers have had to journey to Europe or Japan in order to perform large-scale computations.

To enhance the research of the International Solar-Terrestrial Physics Program, and to promote fruitful interchanges between solar system and astrophysical plasma physics,

We recommend: a national computational program dedicated to basic plasma physics, space

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

physics, and astrophysics, which will provide and maintain state-of-the-art technology appropriate to large-scale theoretical models and simulations. Such a program should ensure ready access to advanced computing on the basis of peer review.

Many problems central to solar-system and astrophysical plasma research are ready for advanced numerical modelling. If these problems are combined with others in hydrodynamics and general astrophysics that should be included in the program, the increased scientific demand, and more importantly, the scientific payoff, would justify the dedicated effort we propose. The success of numerical computations applied to laboratory plasma problems gives us confidence that a computational program would significantly advance space and astrophysical plasma research. Solar-terrestrial research would be the principal initial contributor to, and beneficiary of, such a program. Solar-terrestrial models that successfully meet the test of detailed measurements of both large- and small-scale processes would substantially increase our confidence in models of more distant astrophysical systems.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Appendix A
The Goals of Solar-Terrestrial Physics

The research goals of solar-terrestrial physics have been succinctly defined by the Space Science Board in *Solar-System Space Physics, a Research Strategy* (National Academy of Sciences, 1980).

Two basic principles guided the formulation of these goals:

- The objectives of solar-system space research are to understand the physics of the sun; the heliosphere; and the magnetosphere, ionosphere, and upper atmosphere of the earth, other planets, and comets.
- Studies of the interactive processes that generate solar variation and link it to the earth should be emphasized, because they reveal basic physical mechanisms and have useful applications.

To understand better all the processes linking the solar interior to the corona, we need to study the following:

- The sun's global circulation, how it reflects interior dynamics, could modify luminosity, and is related to the solar cycle.
- The interactions of solar plasma with strong magnetic fields-active regions, sunspots, and fine-scale magnetic knots—and how they cause the release of solar-flare energy to the heliosphere.
- The solar corona's energy sources and the physics of its large-scale weak magnetic field.

To understand better the transport of energy, momentum, energetic particles, plasma, and magnetic field through interplanetary space, we need to study the following:

- First and foremost, the coronal processes that govern the generation, structure, and variability of the solar wind.
- The three-dimensional properties of the solar wind and heliosphere.
- The plasma processes that regulate solar-wind transport and accelerate energetic particles throughout the heliosphere.

To understand better the time-dependent interaction between the solar wind and earth, we need to study the following:

- The transport of energy, momentum, plasma, and magnetic and electric fields across the magnetopause.
- The storage and release of energy in the earth's magnetic tail.
- The origin and fate of the plasma(s) within the magnetosphere.
- How the earth's magnetosphere, ionosphere, and atmosphere interact.

To understand better the entire upper atmosphere as one dynamic, radiating and chemically active fluid, we should study the following:

- The radiant energy balance, chemistry, and dynamics of the mesosphere and stratosphere and their interactions with atmospheric layers above and below.
- The worldwide effects of the magnetosphere's interaction with the polar thermosphere and mesosphere and the role of electric fields in the earth's atmosphere and space environment.
- The effects of variable photon and energetic particle fluxes on the thermosphere and on chemically active minor constituents of the mesophere and stratosphere.

To understand better the effects of the solar cycle, solar activity, and solar-wind disturbances upon earth, we need to:

- Provide to the extent possible simultaneous measurements on many links in the chain of interactions linking solar perturbations to their terrestrial response.
- Create and test increasingly comprehensive quantitative models of these processes.

To clarify the possible solar-terrestrial

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

influence on earth's weather and climate, we need to:

- Determine if variations in solar luminosity and spectral irradiance sufficient to modify weather and climate exist.
- Ascertain whether any processes involving solar and magnetospheric variability can cause measurable changes in the earth's lower atmosphere.
- Strengthen correlation studies of solar-terrestrial, climatological, and meteorological data.

Appendix B

The Unity of Laboratory, Solar System, and Astrophysical Plasma Physics

Although the experimental techniques we use to study laboratory, solar system, and astrophysical plasmas differ, our research on them is unified by the use of plasma concepts. The Committee on Space Physics (National Academy of Sciences, 1984) and the Subpanel on Space and Astrophysical Plasma Physics of the *Survey of Physics* (National Academy of Sciences, in preparation) have concluded that these disciplines share a common set of basic physical problems. These include:

1. magnetic field reconnection;
2. the interaction of turbulence with magnetic fields;
3. the behavior of large-scale plasma flows and their interactions with magnetic and gravitational fields;
4. the acceleration of energetic particles;
5. particle confinement and transport;
6. collisionless shocks;
7. beam-plasma interactions and the generation of electromagnetic radiation; and
8. collective interactions between neutral gases and plasmas.

The fact that such problems emerge from a variety of contexts demonstrates their general significance and suggests that their solution will find further applicability in contexts we cannot imagine today. The existence of such general problems provides a basis upon which a network of common interests, personal interactions, and a common discipline are being built.

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Report of the Research Briefing Panel on Selected Opportunities in Physics

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Research Briefing Panel on Selected Opportunities in Physics

Hans Frauenfelder (*Co-chairman*), University of Illinois, Urbana, Ill.
Mildred S. Dresselhaus (*Co-chairman*), Massachusetts Institute of Technology, Cambridge, Mass.
William F. Brinkman, Sandia National Laboratory, Albuquerque, N. Mex.
Praveen Chaudhari, International Business Machines Corporation, Yorktown Heights, N.Y.
Joseph Cerny, Lawrence Berkeley Laboratory, Berkeley, Calif.
John M. Dawson, University of California at Los Angeles, Los Angeles, Calif.
Harold P. Furth, Princeton University, Princeton, N.J.
Daniel Kleppner, Massachusetts Institute of Technology, Cambridge, Mass.
Alexei A. Maradudin, University of California at Irvine, Irvine, Calif.
Francis M. Pipkin, Harvard University, Cambridge, Mass.
John Schiffer, Argonne National Laboratory, Argonne, Ill.
Watt W. Webb, Cornell University, Ithaca, N.Y.
David Wilkinson, Princeton University, Princeton, N.J.

Staff

Donald C. Shapero, *Staff Director*, Board on Physics and Astronomy, National Research Council
Ellen Williams, *Rapporteur*, University of Maryland, College Park, Md.
Helene Patterson, *Staff Assistant*, Board on Physics and Astronomy, National Research Council
Florence Wong, *Secretary*, Board on Physics and Astronomy, National Research Council
Allan R. Hoffman, *Executive Director*, Committee on Science, Engineering, and Public Policy

The panel gratefully acknowledges assistance in preparing its report from Gordon Baym, Robert B. Palmer, Bruce Partridge, David Schramm, Lee S. Schroeder, Andrew M. Sessler, Charles P. Slichter, and Robert Wagoner. Thanks are also due to participants in a planning meeting to select the topics treated in the Physics Research Briefing; those not named above include Richard Brewer, IBM; Dale Compton, Ford Motor Company; Robert Park, University of Maryland and American Physical Society/Washington; and Robert Richardson, Cornell University.

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

Report of the Research Briefing Panel on Selected Opportunities in Physics

INTRODUCTION

Physics has been rich in discoveries of fundamental laws of nature. These discoveries have influenced other fields of science from mathematics to biology and medicine and have spawned entire industries. Maxwell's electromagnetic theory, for example, forms the basis for understanding electromagnetic phenomena from waves to plasmas and underlies radio, television, radar, and our industrial power networks and modern communication systems. Quantum mechanics provides a framework for portraying physical reality and underlies all of the natural sciences. It has made possible the invention of the transistor, many of the amazing devices of modern solid state electronics, and the laser. Lasers, in turn, are having a major impact on many areas of science. They are also playing a rapidly expanding role in medicine and broad areas of technology, including communications, manufacturing, and national defense.

The discovery of x rays led to the largest single advance in the history of medical diagnostics. The recently created NMR body imaging technique may mark a comparable advance. From the study of electrical noise in the atmosphere, radioastronomy was born, which in turn provided a dazzling new portrait of the universe. Investigations of the properties of nuclei led to the creation of the nuclear power industry. Numerous other examples could be cited. The benefits to society from basic research in physics have been incalculable.

Recent fundamental advances in physics demonstrate that physics is still in a golden age and has never been more vigorous or more productively interactive with other fields. Progress in physics over the last decade has been remarkable. Puzzles that seemed to present insuperable challenges at the beginning of the 1970s have yielded to powerful and elegant theoretical and experimental techniques. The new insights and accomplishments have not only brought greater unity to the various branches of physics but have also strengthened the ties of physics to other areas of science and opened a vast array of new opportunities. Every part of physics has participated in the advance, as the forthcoming report of the Physics Survey Committee will make clear.

The present report focuses on six areas of special opportunity in physics. In order to find areas in which incremental funding may lead

Copyright © National Academy of Sciences. All rights reserved.

to major advances, a planning group (consisting of the Briefing Panel and others listed above) was convened. From twenty suggestions, the following six were chosen: (1) physics at the laser-atomic frontier, (2) relativistic plasma waves, (3) physical properties of deliberately-structured materials, (4) biomolecular dynamics and intercellular cooperativity, (5) cosmology, and (6) nuclear matter under extreme conditions. Large national facilities and well-established programs were explicitly excluded.

The six areas chosen promise to yield fundamental results of great interest. Many of the areas are likely to advance technology and to produce results with impact on the nation's industries; many can be expected to contribute to national security.

All six areas cut across lines of narrow specialization and link different fields. Progress in any one area can thus be expected to stimulate progress in other fields. Cosmology, for instance, connects physics to astronomy and astrophysics. Biological physics interacts strongly with chemistry, biochemistry, biology, medicine, and pharmacology. Deliberately-structured materials relate fundamental aspects of physics to many fields of technology.

1. PHYSICS AT THE LASER-ATOMIC FRONTIER

The laser continues to have a revolutionary impact in the field from which it emerged—atomic physics. The laser-atomic frontier comprises research opportunities generated by joining laser methods with new techniques in atomic and molecular physics.

Opportunities at the laser-atomic frontier include research with trapped particles; femtosecond (10^{-15} sec) spectroscopy and other new forms of spectroscopy; the study of previously inaccessible atomic, molecular, and ionic species; and the development of novel light sources. Advances in these areas will lead to new tests of basic theory and deepen our understanding of the structure of matter at the atomic and molecular level, including the transfer of energy and the nature of chemical reactions. In addition to high scientific interest, many of these topics are likely to yield technological advances. The opportunities selected include the following.

TRAPPED PARTICLES

Using laser light, it is possible to cool to the millikelvin region ions confined in electromagnetic traps and also to cool and trap neutral atoms. In conjunction with recently developed ultrastable tunable lasers, these developments open the way to major advances in high-precision measurements: frequencies to parts in 10^{16}, masses to parts in 10^{11}, new tests of the isotropy of space, and studies of collective motion in plasmas and gases. Applications include advanced atomic clocks, optical frequency standards, metrology, and communications.

NEW SPECTROSCOPIES

Breakthroughs in the generation of femtosecond light pulses make it possible to take "snapshots" of atoms as they collide or molecules as they react, and to observe fast adsorption and relaxation on surfaces. Femtosecond spectroscopy has applications to fast electron circuitry and high-speed instrumentation. Another new type of spectroscopy is carried out with relativistic particle beams. Using the Doppler effect to shift highly stable laser radiation from the visible into the ultraviolet opens the way to new areas of high-resolution spectroscopy of atoms, ions, and molecules. Scientific objectives include studies of relativistic and quantum electrodynamic effects in highly charged ions, high resolution ultraviolet spectroscopy of atoms and molecules, and new types of photoejection measurements.

THE CREATION OF NEW SPECIES

Lasers can produce new species as well as species that have been inaccessible, such as multiply excited atoms, molecular ions, and

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

clusters. Problems such as correlated electron motion, the evolution of matter from single atoms to the condensed state, and catalysis can be investigated. There are many applications to other sciences and to industrial problems. Studies of ionic species can lead to the design of more efficient combustion engines; catalysis is of enormous importance to chemical processing.

Matter in Intense Fields

Laser methods open the way to the study of matter under extraordinary conditions, such as in intense electric and magnetic fields and powerful radiation fields where nonlinear phenomena and multiphoton processes occur. Studies of multiphoton processes made possible by high-power lasers are expected to reveal new aspects of the interaction of photons with matter, and they may open the way to novel types of photochemical processing and isotope separation.

New Light Sources

The creation of the excimer laser and the generation of extreme ultraviolet light in supersonic atomic beams are two recent examples of light sources from research at the laser-atomic frontier. Other new sources can be expected, with wavelengths from the infrared to the soft x-ray region. Such sources would be valuable to wide areas of atomic physics, chemistry, condensed matter physics, and biology. Using coherent generation of pulsed laser light, for example, it may be possible to provide very short pulses in the soft x-ray region. This technique may provide a laboratory-based alternative to synchrotron radiation light sources for many applications.

2. RELATIVISTIC PLASMA WAVES

Recently an exciting area of plasma physics has emerged—that of relativistic plasma waves in which the electrons or both electrons and ions have relativistic velocities. Exploration of relativistic plasma waves is expected to elucidate recently discovered exotic astrophysical objects and the acceleration of particles to create cosmic rays. An understanding of relativistic collective electromagnetic phenomena may lead to novel particle accelerators and radiation generators. We sketch below some of the specific ideas that need investigation.

Particle Accelerators

The next generation of high-energy accelerators is pushing existing technology close to the limit. The long-range future of particle and nuclear physics may well depend on the development of new acceleration techniques. In addition, new accelerators in the range of 10 MeV to 10 GeV may find application in physics, other sciences, industry, and medicine.

Plasma Accelerators

Space-charge waves in plasmas can produce electric fields that are orders of magnitude larger than can be generated by any other means, with the possible exception of a laser beam focused in a vacuum. These fields can be larger than the fields that bind electrons to atoms (10^9 volts/cm). Such fields can only be maintained in a plasma because they instantly ionize ordinary materials. In a beat-wave accelerator, space-charge waves are generated by sending through the plasma two intense laser beams with a frequency difference equal to the plasma frequency. The combination of high electric fields and phase velocities close to that of light makes such waves good candidates for the acceleration of particles to high energy.

Grating Accelerators

Another novel method uses a grating of droplets illuminated by a powerful short-pulse laser. The resulting plasma grating generates fields estimated to be of several hundred million volts per cm.

Copyright © National Academy of Sciences. All rights reserved.

Free-Electron Laser Accelerators

Another possibility for constructing an accelerator is to operate a free electron laser "backwards" in an accelerating mode. A further possibility is to operate a free electron laser "forward" and use the resulting radiation to accelerate particles in a conventional linear accelerator. Both of these concepts pose interesting problems and will require developments in current technology.

Laser Development

All of these accelerating schemes require terawatt lasers with good optical quality and very short pulses (shorter than 1 picosecond). Progress with these approaches could be accelerated by a laser development program aimed toward these requirements.

Generation of Electromagnetic Radiation

The inverse of particle acceleration—radiation by energetic particles in plasmas—is a challenging subject of scientific and potentially practical interest. Bursts of radiation that accompany solar activity and are also emitted from a wide variety of astrophysical objects are poorly understood. Recent observations of radiation from pulsars show that the pulses have γ-ray energies up to 10^{15} eV. Such energies imply that the acceleration process occurs over relatively short distances (of the order of 10^3 km) with accelerating fields larger than 10^7 volts/cm. Relativistic waves may be involved. Closer to home, bursts of energetic particles are generated by solar flares; here again strong electric fields must exist in the plasma. The theoretical and computer modeling tools are now reaching the point where progress in understanding these processes can be made.

In the laboratory, copious radiation, roughly up to the thirtieth harmonic of the plasma frequency, has been observed when relativistic electron beams are injected into plasmas. The generation of electromagnetic radiation by means of free-electron lasers is actively pursued at a number of laboratories. Within the last year, free-electron lasers have started operation—two in the microwave, three in the infrared, and one in the visible range. Further configurations especially suited to different parts of the spectrum could be developed. Moreover, theory predicts that energetic electrons streaming through plasmas with strong density fluctuations should yield radiation similar to that produced in present free electron lasers where rippled magnetic fields are used instead of density ripples. At present there are virtually no controlled experiments to investigate these phenomena. The computational tools to address these problems exist, but the resources for using these tools are generally not available to university researchers.

3. DELIBERATELY-STRUCTURED MATERIALS

Condensed-matter science has a long and continuing history of contributions to basic understanding of physical phenomena. General concepts from this field are bringing greater intellectual unity to physics. Ideas from condensed-matter physics and many-body theory are playing a role in such diverse fields as nuclear physics, astrophysics, and theoretical methods of particle physics. While condensed-matter physics is intellectually a rich and fertile field of inquiry for the basic questions of science, it is also playing a vital role in advanced technology from microelectronics and communications to energy systems.

From the vast range of activities in condensed-matter physics, we emphasize only one area of special opportunity: deliberately-structured materials. By deliberately-structured materials, we mean materials designed by man to have particular structures. Their fabrication may involve controlling on an atomic length scale the compositional or

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

structural arrangements of atoms in one or two dimensions. The resulting materials possess remarkable properties not otherwise found in nature—properties that, in many instances, present major scientific puzzles.

Not only are these materials likely to raise fundamental questions of physics, but their study will also reinforce the technological base of the nation's most important device-related industries. Conversely, technologies that have led to advances in the fabrication of very-large-scale integrated circuits can be harnessed to create deliberately-structured materials.

Three research areas that are currently of great scientific interest and that are especially likely to advance through the use of deliberately-structured materials are: (1) studies of surfaces and interfaces between specifically designed microscopic structures; (2) investigations of "disorder" in matter that might be understood more deeply through tailoring of its composition or geometrical structure; and (3) studies of novel collective phenomena in "exotic materials."

Surfaces, Interfaces, and Thin Films

Beginning with an atomically clean and crystallographically defined surface, it is possible to prepare a two-dimensional film only 1 atom thick by deposition of a second material. The film can be atomically smooth or contain crystallographic discontinuities, depending on the structure of the substrate. Its structure can be further modified by pattern etching on a scale finer than 100 Å. Deposition of additional layers of dopants, or of metals, semiconductors, superconductors, or insulators, leads to atomically sharp interfaces of defined structure and composition. The physics of these low-dimensional systems is fundamentally interesting. Furthermore, studies of the electronic, vibrational, magnetic, and transport properties of interfaces are applicable to the technological needs of the computer and energy industries.

Disorder in Matter

There are many approaches to creating materials with controlled disorder. One method is the random dispersion of very small particles of one material in a matrix of a second. Another is the fabrication of alternating amorphous layers, resulting in two-dimensional disorder within the plane of the layers but one-dimensional order perpendicular to the plane due to the stacking of the layers. Spin glasses present a totally different form of disorder in which the atoms may be compositionally ordered but the spin-spin interactions are random. Metastable (i.e., nonequilibrium) disordered structures represent a fourth class of deliberately-disordered materials. All these materials have properties that are intrinsically interesting and important. In addition, insight into the nature of controlled disorder may provide a basis for understanding technologically important disordered materials such as polymers, glasses, and composite materials and may even lead to a better understanding of problems in computation, optimization, and neurobiology.

Exotic Materials

The structure of materials can be arranged in many different ways to produce unique physical properties. For instance, it is possible to deliberately construct ordered crystals in arrays of alternating layers called superlattices. The properties of superlattices can be modified in desirable ways by tailoring the compositional and impurity profiles of the constituent layers. Many of the traditional methods of condensed-matter theory must be extended to account for the small size or the lower dimensionality of such systems. New phenomena, undreamt of a short time ago, have been discovered in deliberately-structured materials. Examples include quantization and fractional quantization of the Hall effect. The same class of materials that is actively used to investigate the quantum Hall

Copyright © National Academy of Sciences. All rights reserved.

effect is also being examined by researchers worldwide for new applications in the computer and communication industries.

Needs of the Field

The overwhelming characteristic of this field is its dependence on a strong infrastructure to support a range of activities from materials preparation and characterization to study of the nature and causes of unusual materials properties. The infrastructure, involving skilled technicians and complex, state-of-the-art apparatus and instruments, does not exist in most universities; and the institutional mechanisms to put the structure in place at universities are often lacking. The intellectual concepts are there to be exploited if the university infrastructure can be enhanced.

4. BIOMOLECULAR DYNAMICS AND INTERCELLULAR COOPERATIVITY

The magnificent complexity of life that we study as biology ultimately reflects underlying physical principles. At each level—biomolecules, supramolecular structures, cells, and organisms—the behavior of the system is governed by physical laws. The detailed description of biological phenomena in terms of physical laws promises to yield rich dividends both in terms of understanding life more deeply and in exploring the physics of complex systems. Research on biological systems connects physics to chemistry, biochemistry, biology, and, ultimately, pharmacology and medicine. The systems under investigation are, of course, the same in all these fields, but the approaches to them are different. The physicist selects the simplest available system, performs quantitative measurements, constructs a model, and attempts to describe the observed behavior fully in terms of known physical laws. The following examples, at different levels of complexity, indicate where this approach is beginning to bear fruit.

Biomolecular Dynamics

Proteins and nucleic acids, the building blocks of life, were called "aperiodic crystals" by Schrödinger. Seen by standard x-ray diffraction, they indeed appear to be like crystals with linear dimensions of a few nanometers. In recent years, however, a far more exciting picture has emerged. Biomolecules are dynamic systems that represent a state of matter different from solids, liquids, and gases. They show many similarities to glasses and spin glasses but are far more sophisticated. The sophistication is not surprising; each biomolecule has undergone more than 10^9 years of research and development and is built for a particular function. A protein, for instance, must have precise mechanical, electrical, thermal, and chemical properties. These properties are not spatially homogeneous but rather vary throughout the protein volume. Through genetic engineering, proteins with atomic modifications in desired places can be produced. The entire field of physical studies of biomolecules is at its very beginning, but the importance and the potential of such studies are increasingly evident. Biomolecules are excellent laboratories for the investigation of basic physical laws. We can study elementary excitations, motions, transport phenomena, and elementary reaction steps. Biomolecular physics overlaps the physics of glasses and spin glasses and complements the chemistry of reactivity, catalysis, and life processes. With increasing knowledge of the relation between structure and function, it will become possible to design and synthesize biomolecules with desired properties for uses in biological research, medicine, and pharmacology. It is also likely that an understanding of the relation between structure and function of biomolecules, together with genetic engineering, will produce spectacular results in other areas. Biosensors have the potential of improving equipment from medicine to information transfer. New bio-

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

materials, such as biomolecular catalysts and artificial photosynthetic systems, may find uses in areas ranging from agriculture to space.

Transmembrane Signalling

Transmembrane signalling is conspicuous in brain, nerve, and muscle tissue. Its molecular basis is now directly accessible through a new biophysical technology: recording of the stochastic electrical conductances of individual molecular channels isolated in very small patches of cell membranes or reconstituted into model membranes. Switching of transmembrane currents through molecular channels is controlled by neurotransmitters, messenger molecules, or membrane potential changes. Hundreds of distinguishable membrane channels and nearly as many neurotransmitters have been discovered. How are channel protein conductances switched by binding of a neurotransmitter? How is ion selectivity built into the protein structure? These profound questions and the biological behavior and disorders associated with them appear susceptible to solution with the help of a combination of gene manipulation and biophysical measurements.

Molecular Basis of Information Storage

Elegant experiments on the organization of elementary learning and memory in simple animals at the cellular and molecular level imply a variety of basic signalling processes. The molecular basis of memory is, however, still not understood. The combination of advanced biophysical techniques, such as sensitive optical and electron microscopy probes, with hybridoma techniques for producing monoclonal antibodies, offers hope of elucidating the molecular features of information storage and signal transmission. Prospects are good for understanding the molecular mechanisms of auditory transduction and tone discrimination and the mechanism of visual signal transmission from the photoreceptor. A new family of sensors may develop from these biophysical experiments with some hope of revolutionizing the tools of clinical medicine.

Intercellular Cooperative Processes

Recent developments at the interface between physics and mathematics have generated new analytical concepts dealing with discrete and nonlinear systems. These systems often display regimes of solution characterized by solitons, limit cycles, disorder, and chaos. These concepts represent many intercellular cooperative processes and have provided understanding of several systems, such as the cooperative heart muscle cell contraction in the heart beat and its failure in ventricular fibrillation.

Physicists have generated new conceptual models of distributed memory. These models, while speculative at present, may, if successful, find ultimate applications in robotics and computer architecture. The theory reproduces many of the foibles and perhaps some of the strengths of the human memory.

5. COSMOLOGY

Few areas of science enjoy as wide an appeal to scientists and laypersons as cosmology. Fascination with the origin and fate of the universe is deeply imbedded in the human intellect, and speculations about these ultimate questions have stimulated some of the most profound thinking in modern science. As in all other scientific disciplines, the basis of cosmology lies in experimental observations. Accurate measurements of the abundances of light elements, for example, fit the specific predictions of a "hot big-bang" model for our universe. Light nuclei were formed when the universe was a few minutes old and had a temperature of 10^9 K—a fantastic extrapolation of known physics. The same hot

Copyright © National Academy of Sciences. All rights reserved.

big-bang model predicted the 3 K cosmic radiation a decade before its discovery.

Cosmology and Particle Physics

Though uniquely successful in explaining these fundamental observations, the hot big-bang model by itself fails to explain others.

- Why is our universe so asymmetric in matter and antimatter?
- Why is the ratio of baryons to photons in the universe $\sim 3 \times 10^{-10}$?
- Why is the kinetic energy in universal expansion so nearly equal to the universe's gravitational potential energy at the current epoch?
- How did the universe reach the same temperature to a part in 10^4 at points not causally connected?

Two recent ideas from particle physics may answer these fundamental questions and still fit into the basic hot big-bang picture. Grand Unified Theories (GUTs) when applied in the very early universe (age $\sim 10^{-35}$ sec) supply answers to the first two questions. Thus, laboratory tests of GUTs (e.g., by means of proton-decay experiments) may also resolve important cosmological problems. The last two questions above may be answered by a model in which the universe expanded exponentially for a brief time early in its history. This "inflation" is supposed to have been driven by the vacuum energy of a scalar field perhaps related to the Higgs particle that is required by the new theories of electromagnetic and weak interactions. We can, of course, turn the arguments around. Already cosmological data are constraining and guiding GUTs, making the early universe effectively a particle physics laboratory.

What Lies Ahead?

The long-awaited evidence to describe the fate of the universe could come from either of two lines of attack. One approach attempts to measure directly the deceleration of the universe over the past few billion years. The other approach looks for "invisible" mass density that might eventually reverse the current universal expansion, leading to a hot recollapse. Massive neutrinos, axions, and other exotic particles are under active study in this context by particle physicists and cosmologists. Primordial nucleosynthesis has recently advanced a strong argument against the existence of a large component of dark matter in baryonic form; thus, the search for an exotic mass component has intensified.

Another area of current excitement is the origin and development of large-scale structure—galaxies, clusters of galaxies, etc. The structure seen today is assumed to grow from density perturbations in the early universe. However, observational limits on fluctuations in the thermal background radiation at small angular scale severely constrain models for the origin of structure. Most models that lead to the observed matter clumping also predict fluctuations of the radiation exceeding observed limits. The standard picture must be modified and extended.

Cosmology is at present ripe with opportunity and certain to be full of surprises. The combination of the space program with major advances in astronomical instrumentation is rapidly pushing our horizons deeper into the universe across a wide range of the electromagnetic spectrum. Theory, drawing on all areas of physics, is creating a plausible and testable cosmological model rich with beautiful physics and unanswered questions.

6. NUCLEAR MATTER UNDER EXTREME CONDITIONS

Until recently the study of nuclear matter was restricted to normal values of density and low temperature. With the acceleration of heavy nuclei to relativistic energies, an important new dimension has been added to these studies, namely, the ability to deliver unprecedented amounts of energy and momentum into the volume of the colliding nuclei, a volume large relative to that of a

Copyright © National Academy of Sciences. All rights reserved.

Research Briefings, 1984
http://www.nap.edu/catalog.php?record_id=19410

nucleon. With this tool a new era for studying nuclear matter opens up; for the first time we can test the response of such matter to extreme conditions of energy density and compression.

Scientific Opportunities

Central nucleus-nucleus collisions at high energies open up new avenues that cross the traditional boundaries of particle, nuclear, and astrophysics. The broad goals of such a program—to explore the cooperative behavior in extended hadronic systems with many interacting degrees of freedom—are complementary to those of particle physics in the same sense that condensed-matter physics, with its varied and often unexpected phenomena, transcends the underlying atomic physics. The salient difference is that nuclear studies can themselves shed light on the incompletely understood mechanism of confinement in the basic quantum chromodynamic theory.

Foremost among the opportunities is the possibility of producing a new state of matter called the quark-gluon plasma. Through numerical lattice-gauge calculations, quantum chromodynamics predicts that confined hadronic matter, in which quarks are bound together to form baryons and mesons within the nucleus, will cease to exist at sufficiently high values of compression or energy density. A new phase of matter appears in which the quarks and gluons are freed from the confines of the hadrons. Current theoretical estimates and extrapolations of existing data indicate that the phase transition can be explored in at least two distinct domains accessible in heavy-ion collisions. The domain of *maximum nuclear compression* is a baryon-rich environment expected to occur at energies of a few GeV/nucleon in the center-of-mass frame. In this domain, the heaviest nuclei can still stop each other completely. Densities up to 10 times that of normal nuclei should occur; such densities may be achieved through enhancement of existing accelerators. The domain of *high-energy density* is expected to occur in the central rapidity region left after the colliding nuclei pass through each other at energies of several tens of GeV/nucleon in the center-of-mass frame. Exploration of this domain, expected to be meson-rich with energy densities in excess of 10 times those of ordinary nuclei, awaits construction of future accelerators.

Colliding very heavy nuclei at relativistic energies allows one to study new states of matter and to map out the equation of state of highly compressed/excited nuclear matter, while providing the first indications of the dynamical behavior of extended regions of matter under the most extreme baryon and energy densities accessible in the laboratory.

Suitable facilities will soon be available for accelerating light ions to the 15-225 GeV/nucleon range. These new capabilities are in the region expected to yield maximum nuclear compression and to provide an important extension of our studies of nuclear matter under extreme conditions.

By their very nature, these complex experiments present a major challenge to experimentalists. Detector requirements are similar to those found in particle physics but with a high degree of detector segmentation and enhanced particle identification because of the large numbers of particles expected. Research and development efforts on these detectors are needed so that we will be adequately prepared to measure the complex event structures expected in these interactions.

In nuclear theory, substantial efforts are required to produce theoretical benchmarks for comparison with experimental results. Examples include large Monte Carlo intranuclear cascade calculations and lattice-gauge calculations with inclusion of finite quark masses and finite baryon densities.

Impact on Other Areas

In addition to potential technological spinoffs from the development of new detectors, research on central high-energy nucleus-

Copyright © National Academy of Sciences. All rights reserved.

nucleus collisions links nuclear physics to other fields. Information gained about the transition from the quark-gluon plasma to hadronic matter is important to cosmologists for understanding how the present state and distribution of matter in the universe evolved. Data from the regime of maximum compression are required to understand the state of matter in the deep interior of neutron stars. There is no other way such information on the high-density equation of state and other properties of matter vital for condensed-matter astrophysics can be obtained in the laboratory. These data will facilitate the interpretation of observations from future orbiting x-ray telescopes such as the AXAF and will help to unravel questions about the birth and evolution of neutron stars. Particle physicists have been studying the short-range behavior of quantum chromodynamics and processes of quark confinement; high-energy nuclear collisions will allow us for the first time to investigate the long-range behavior of this theory and the processes of deconfinement.

CONCLUSIONS

The six areas selected in the present report comprise only a small part of physics. But they exhibit clearly the characteristics of physics—enormous diversity, the search for fundamental laws, strong connections to many other sciences, and technological and industrial applications. Each of the six areas is at a stage where incremental funding may produce major progress. The reasons for the high leverage vary from one area to another, but in each area, there could be rapid progress through meeting a need either for seed funding to promote new research or for new instrumentation, especially medium-cost instrumentation for university research laboratories. Several of the subjects—relativistic plasma accelerators, cosmology, and parts of biophysics—do not have established mechanisms for support and review.

The new areas within nuclear physics, biophysics, and cosmology described above are each in their infancy; exploiting the opportunities to do new forefront physics addressing fundamental questions about matter, life, and the universe could be greatly accelerated by a modest investment of seed money.

Each of the new areas within atomic physics, plasma physics, and condensed-matter physics described above has a need for medium-cost instrumentation that is difficult for university laboratories to obtain within existing funding frameworks. An increment of funding targeted to the instrumentation problem in these areas could unleash untapped scientific productivity in areas with potential applications to technology and national security.

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.