SCIENCE

Published by the **American Association for the Advancement of Science (AAAS)**, *Science* serves its readers as a forum for the presentation and discussion of important issues related to the advancement of science, including the presentation of minority or conflicting points of view, rather than by publishing only material on which a consensus has been reached. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by the AAAS or the institutions with which the authors are affiliated.

Membership/Circulation

Director: Michael Spinella
Fulfillment: Marlene Zendell, *Manager*; Gwen Huddle, *Assistant Manager*; Mary Curry, *Member Service Supervisor*; Pat Butler, Helen Williams, Laurie Baker, *Member Service Representatives*
Promotions: Dee Valencia, *Manager*; Hilary Baar, *Coordinator*
Research: Kathleen Markey, *Manager*; Robert Smariga, *Assistant*
Financial Analyst: Jacquelyn Roberts
Administrative Assistant: Nina Araujo de Kobes

Advertising and Finance

Associate Publisher: Beth Rosner
Advertising Sales Manager: Susan A. Meredith
Recruitment Advertising Manager: Janis Crowley
Advertising Business Manager: Deborah Rivera-Wienhold,
Financial: Julie Eastland, *Manager*; Andrew Joyce, *Analyst*
Marketing Manager: Laurie Hallowell
Traffic Manager: Tina Turano
Recruitment: Michele Pearl, *Operations Manager*; Dan Moran, *Traffic Manager*; Debbie Cummings, Millie Muñoz-Cumming, Angela Wheeler, *Sales*
Reprints Manager: Corrine Harris
Permissions Manager: Arlene Ennis
Advertising Assistants: Allison Pritchard, Kelly Nickerson
Send materials to *Science* Advertising, 1333 H Street, NW, Washington, DC 20005, or FAX 202-682-0816.

SALES: Northeast/E. Canada: Fred Dieffenbach, 802-867-5581, FAX 802-867-4464 • **Mid-Atlantic:** Richard Teeling, 201-904-9774, FAX 201-904-9701 • **Southeast:** Mark Anderson, 305-856-8567, FAX 305-856-1056 • **Midwest:** Donald Holbrook, 708-386-6921, FAX 708-386-6950 • **West Coast/W. Canada:** Neil Boylan, 415-673-9265, FAX 415-673-9267 • **Germany/Switzerland/Austria:** Ric Bessford, World Media Services, Germany; +49-089-39-00-55, FAX +49-089-39-00-15 • **Japan and Far East:** Mashy Yoshikawa, Orient Echo, Inc., Japan; +3 3235-5961, FAX +3 3235-5852 • **UK, Scandinavia, France, Italy, Belgium, The Netherlands:** Andrew Davies, Great Britain; +44-457-838-519, FAX +44-457-838-898
European Recruitment: AnneMarie Vis; +44-223-424-695, FAX +44-223-424-695 • **Other:** For recruitment advertising inquiries contact *Science* Advertising: 202-326-6555; For product advertising inquiries contact 202-326-6544, FAX 202-682-0816.

■

Information for Contributors appears on pages 600-602 of the 31 July 1992 issue. Editorial correspondence, including requests for permission to reprint and reprint orders, should be sent to 1333 H Street, NW, Washington, DC 20005. **Science Telephone:** 202-326-6500, TDD 202-408-7770. London office: 071-435-4291. **Subscription/Member Benefits Questions:** 202-326-6417. **Other AAAS Programs:** 202-326-6400.

LETTERS

Computer Speed and Sequence Comparison

Despite recent well-known advances in computer performance, it is still a commonly and erroneously held belief that rigorous sequence comparison methods are too expensive to use for protein database scanning.

In their recent report "Exhaustive matching of the entire protein sequence database," (5 June, p. 1443), Gaston H. Gonnet *et al.* suggest that the full self-comparison of the protein database would require 10^6 years of computer time (*1*) without reorganization of the sequence data by indexing on a patricia tree. The implication is that such a task is beyond the capabilities of today's workstations.

The Smith-Waterman (*2*) local similarity algorithm, in common with other rigorous methods (*1*, *3*), requires *MN* steps to calculate the optimum score for aligning two sequences of length M and *N*, which includes a consideration of insertions and deletions. To provide a fair test of modern computers, I have implemented this algorithm in C language, taking care to optimize the most frequently executed parts of the program.

When run on a Hewlett-Packard 730 (HP730) workstation, with the 141-residue human α-hemoglobin as a query, the program takes 368 seconds to scan the 25,044 sequences (8,375,696 residues) in the SwissProt database release number 22. Six minutes is a reasonable scan time that compares favorably with IntelliGenetics Inc.'s BLAZE, one of the speediest programs. For the same query, BLAZE runs a factor of 17.5 faster, but only on a dedicated 4096 processor MasPar computer. The HP730 time corresponds to 3.2 million array operations per second. The complete, rigorous self-comparison of the SwissProt data that has been considered impossible would require 3.5×10^{13} such operations, or 4.5 months of central processing unit time to complete on an HP730. Simple distributed processing techniques could reduce this figure to a few weeks (*5*). Single processor computers from DEC promise speeds up to four times the HP730 and would provide similar times on a single workstation.

Since the mid-1980s, the speed of typical laboratory computers has increased by a factor of 70, while their cost has been reduced by a factor of 10. In contrast, over the same period the database of known protein sequences has increased only by a factor of 8. If we ignore cost-performance gains, today's conventional computers are at least nine times faster at scanning today's database than the machines in 1985 were on a contemporary databank. It seems likely that single processor computer technology will continue to keep ahead of the protein database until the large-scale automation of DNA sequencing becomes a reality.

Geoffrey J. Barton
Laboratory of Molecular Biophysics, South Parks Road, Oxford OX1 3QU, United Kingdom

REFERENCES

1. S. B. Needleman and C. D. Wunsch, *J. Mol. Biol.* 48, 443 (1970).
2. T. F. Smith and M. S. Waterman, *ibid.* 147, 195 (1981).
3. P. H. Sellers, *SIAM J. Appl. Math.* 26, 787 (1974).
4. G. J. Barton, *Comput. Appl. Biosci.* 7, 85 (1991).

Response: We stated in our article that "an exhaustive matching of the entire sequence database with the Needleman-Wunsch (*1*) algorithm *has been thought to be* essentially impossible" (italics added), quoting an expert in the field (*2*). If each pairwise comparison required 1 second to execute, *then* the time required for an algorithmically naive matching would be greater than a million years.

To the extent that we (and others) have not fully recognized the computational power of some of today's fastest computers, Barton's comments are enlightening. However, our approach and Barton's differ in three fundamental ways, all interesting to the general scientist who wants to use sequence alignments without becoming entangled in its mathematics.

First, there are always two approaches to solving major computational problems. One relies on ever-increasing computational power to speed brute force solutions. Our approach is to improve the algorithms used. While both improved computers and improved algorithms will speed any process, Barton's assured reliance on increasing computational power alone is based on a misleading calculation about the ability of such power (up by a factor or 70 since the mid-1980s) to keep pace with growth in the database (up by a factor of 8). As an exhaustive matching by naive algorithms increases with the size of the data base *squared*, computational power has in fact barely kept pace with increase in the size of the database, even without genome sequencing projects and automated sequence collection.

Next, Barton implements the Smith-Waterman (SW) algorithm (*3*), which finds only a single match between two sequences. For example, let sequence A be composed of

blocks [*a b c d*] and sequence B of blocks [*w x y z*]. Suppose that the best alignment of portions of the two sequences is *b*:*x*. SW will report it. However, other close (but less close) matches (for example, *z*:*d*) will be overlooked by SW. Further, the SW algorithm does not detect self-matches within a sequence. Thus, one reason Barton's approach is so fast is because it does not yield an exhaustive matching.

Finally, the practicing biochemists among us are always impressed by the power of modern computers. We are more impressed, however, when this power is applied to solve interesting problems. Predicting de novo the folded structure of proteins (*4*), designing de novo proteins that fold and catalyze reactions (*5*), and constructing models for ancient forms of life (*6*) are activities that can benefit from our exhaustive matching. Perhaps the most important point of Barton's note is that neither these activities nor the exhaustive matching itself came from those using the most advanced computers.

Steven A. Benner
Mark A. Cohen
Institute for Organic Chemistry,
Eldgenössische Technische Hochschule,
8092 Zurich, Switzerland
Gaston H. Gonnet
Institute for Scientific Computation,
Eldgenössische Technische Hochschule

REFERENCES

1. S. B. Needleman and C. D. Wunsch, *J. Mol. Biol.* **48**, 443 (1970).
2. W. Taylor, *Nature* **353**, 388 (1991).
3. T. F. Smith and M. S. Waterman *J. Mol. Biol.* **147**, 195 (1981).
4. S. A. Benner, *Curr. Opin. Struct. Biol.* **2**, 402 (1992).
5. K. Johnsson, R. K. Allemann, S. A. Benner, in *Molecular Mechanisms in Bioorganic Processes*, C. Bleasdale and B. T. Golding, Eds. (Royal Society of Chemistry, London, United Kingdom, 1990), p. 166.
6. S. A. Benner, A. D. Ellington, A. Tauer, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7054 (1989).

■

Yeast Biology

I read with interest the Research News article "Yeast biology enters a surprising new phase" by Michelle Hoffman (20 Mar., p. 1510). The article states that "*Saccharomyces* was believed to have only the unicellular phase. Until now." The authors of the *Cell* paper discussed in the article (*1*) describe pseudohyphal growth, a dimorphic transition in the life cycle of *Saccharomyces cerevisiae*. Although they refer to earlier work in which pseudohyphal growth was noted (Guilliermond, 1920; Brown and Hough, 1965; Lodder, 1970; Eubanks and Beuchat, 1982), they classify these as anecdotal references. It is surprising to me that in the *Cell* article no mention was made of the work done by Lindegren and others on the genetics of yeast in the 1950s. Townsend and Lindegren (*2*) described growth patterns of different members of a polyploid series of *Saccharomyces*. There are also any number of botany textbooks that refer to the pseudohyphae in *Saccharomyces*. Dittmer's text *Phylogeny and Form in the Plant Kingdom* (*3*) states, "in a colony of actively growing yeast it is quite common to find chains of cells with hyphalike characteristics."

R. Witkus
Department of Biological Sciences,
Larkin Hall,
Fordham University,
Bronx, NY 10458

REFERENCES

1. C. J. Gimeno, P. O. Ljungdahl, C. A. Styles, G. Fink, *Cell* **68**, 1077 (1992).
2. G. F. Townsend and C. C. Lindegren, *J. Bacteriol.* **67**, 480 (1954).
3. H. J. Dittmer, *Phylogeny and Form in the Plant Kingdom* (Van Nostrand, New York, 1964).

In contrast to the statements in the article to this effect, it has been known for more than 70 years that *Saccharomyces cerevisiae*, among many other yeasts, may form pseudomycelium under special conditions. In fact, this property is a major determination characteristic in yeast taxonomy and identification. The phenomenon is well documented for *Saccharomyces* species (*1*). The media used to obtain pseudomycelium reproducibly (or true mycelium for other yeasts) have been described in detail in the taxonomic handbooks. While we agree that a molecular study of the formation of pseudomycelia in *Saccharomyces cerevisae* has not been carried out before, we stress that the occurrence of pseudomycelium in baker's yeast is well known to almost any yeast physiologist or taxonomist.

H. Y. Steensma
Department of Cellbiology and Genetics,
Leiden University,
Leiden, The Netherlands
Maudy T. Smith
Centraalbureau voor Schimmelcultures,
Yeast Division,
Delft, The Netherlands

REFERENCES

1. J. P. van der Walt, in *The Yeasts*, J. Lodder, Ed. (North-Holland, Amsterdam, 1970), p. 599; J. A. Barnett, R. W. Payne, D. Yarrow, *Yeasts* (Cambridge Univ. Press, Cambridge, UK, 1990).

Hoffman writes that "[w]hile many related molds, such as the human pathogen *Candida albicans*, exist in two phases (a unicellular yeast and a multicellular filamentous phase) . . . *Saccharomyces* was believed to have only the unicellular yeast phase. Until now." The article proceeds to elaborate on the "news" that *Saccharomyces* can also produce a filamentous phase.

In 1886, E. C. Hansen described *Saccharomyces cerevisiae* Hansen (*1*), and in 1890, W. Zopf wrote (*2*) (Fig. 1) (roughly translated)

E. Chr. Hansen had first reported that the Saccharomyces-like yeasts, in general, are capable of producing another shape, that is, a typical chain-like mycelium. This is found in a particularly distinct form in the beer yeasts, for example, in *Saccharomyces cerevisiae* Hansen. From this it is clear that the view found in all books, that the yeast fungi are "one-celled" growth forms, is to be dismissed as erroneous.

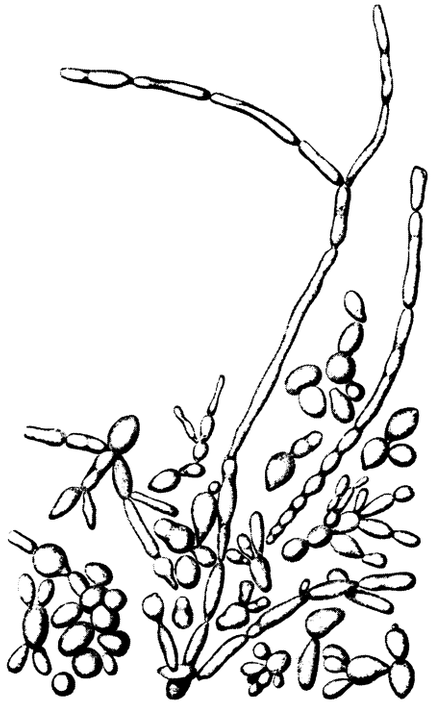

Fig. 1. Drawing of *Saccharomyces cerevisiae* from Hansen's description, reproduced by Zopf (*2*), showing the filamentous growth of the species. [From (*2*), figure 113, p. 421)]

Elsewhere in Hoffman's article and in the original *Cell* paper by Gimeno *et al.* are aspects of the yeast-filament switch that are truly newsworthy. Gimeno *et al.* went far beyond Zopf in considering the switch from single-cell to filamentous growth form in relation to ploidy, in looking at polar versus nonpolar growth, and in considering its potential infuture investigations into developmental biology.

Mary L. Berbee
Barbara Kloeckener
John W. Taylor
Department of Plant Biology,
University of California,
Berkeley, CA 94720