THE ROCKEFELLER UNIVERSITY

pro bono humani generis

1230 YORK AVENUE - NEW YORK, NEW YORK 10021-6399

Joshua Lederberg
UNIVERSITY PROFESSOR

June 23, 1990

Edward Shortliffe, M.D., Ph.D.
Professor of Medicine and Computer Science
Stanford University Medical Center
Medical School Office Building, X215
Stanford, Calif. 94305-5479.

Dear Ted:

I am very pleased to have the opportunity to rejoin the old SUMEX team icw your new proposal re the Beckman Electronic Library System. It resonates with my needs and interests in several ways; above all the Beckman Center for Molecular Genetics is an excellent testbed for developing and demonstrating what the new technologies can do for more rapid progress in biomedical research.

This came at a most propitious time for me. In a few days, I am relieved of the administrative responsibilities I have held since I left Stanford 12 years ago, in order to return to a research career at the Rockefeller University. In fact, my prospective program just bridges the computer science and the laboratory aspects of your own proposal; so there is extraordinary economy in my joining forces with you rather than initiating an independent effort.

During the past year, I have been preoccupied with theoretical examination of what is meant by "spontaneity" of mutation, (1) and this has led in turn to an experimental program to test the (probable!) correlation of a gene's transcriptional activity and its vulnerability to mutagenesis and DNA repair (2). I expect to start laboratory work on this subject within the next few months, as quickly as I can renovate my lab facilities and complete recruitment of a small staff. The theoretical work involved acquiring and reading an enormous amount of literature; every day I would discover some new nugget, and some better approach to retrieval keys, to open up further facets of the problem. I also found myself in a formal planning mode, as a harbinger of return to the MOLGEN studies that Peter Friedland and Mark Stefik worked on while I was directing SUMEX. I would like to build on that base, to see how far the conceptual structure of molecular biology can be formalized, so as to provide aids to a) theory formation, b) proof and validation, c) experiment planning, and d) codification of and access to the existing literature. This is all still in a rather primitive stage; and I would have found myself both the initiator and the testbed of any system that was developed -- I much prefer the larger scope of a collaboration with you.

I am very much interested in concept-based research tools, and will make full use of those provided by ISI (citation-based) and ADS. In fact, I am already an intense user of ISI's Science Citation Index on CD-ROM -- it would be a 10-fold enhancement of efficiency and vital interactivity to have real time access to the current articles themselves. And how much time I would save NOT having to download (e.g. xerocopy) the full texts to assure prompt access the next time around. So I fully expect to be the most avid user of your system!

My own approach to concept-search is to go beyond the automated term- and bibliographic coupling, towards a logical analysis of prevailing concepts, and the development of a more formal language of expression. The bottom-up aspects of this have been started in MOLGEN; I have also been starting to work top-down, with particular emphasis on high level concepts whose flaws and insufficiencies have already been brought to historic attention. These "stepping stones turned stumbling blocks" have the advantage of having received a great deal of critical and experimental attention, and therefore an ample record in the literature. They also illustrate the cautions we must exercise before crystallizing any set of rules (ESPECIALLY high level ones!) into expert systems. Examples: "Proteins are Enzymes" -- has been corrected with the discovery of ribozymes. "DNA => RNA" needed to be amplified to account for reverse transcription. "The 3-D folded conformation of a protein is determined by its primary amino-acid sequence" neglects the role of ligands in allosterism, and possibly in dynamic control of folding kinetics, as well as of proline isomerases. I am hoping to develop an automated "logical assistant" that will assist in the dissection of precise meaning of such theoretical assertions, and of the experimental proofs that have been asserted on their behalf. The resources of the Electronic Library Project will be invaluable to me at several levels, from its computer science and methodology to the actual delivery of information. In turn, I hope to be of assistance both in the design and methodological stages, and in evaluation as an ardent user.

For some years, I have sustained an appointment as a Consulting Professor in the Stanford Genetics Department (which I had chaired for 20.) In that capacity, I will be part of the Stanford Team, with recurrent personal visits; and also in a well-exercised and familiar use of network communications -- the fruit of the SUMEX-AIM system that I had pioneered. Outside of my own small laboratory group, I have not as yet made any plans for the direct engagement of other Rockefeller University faculty; if problems of licensing can be surmounted, I would be eager to bring the R.U. library and many of my faculty colleagues into the same orbit. But it would be an undue complication to attempt to formalize that at this time. The R.U. does have the advantage that the entire university maps congruently onto the scientific interests of the Beckman Center.

Yours sincerely,

Joshua Lederberg

1) Lederberg, J. 1989
 Replica plating and indirect selection of bacterial mutants;
 Isolation of preadaptive mutants in bacteria by sib selection.
  Genetics 121:395-399.

2) Further example just came across my desk today: Venema,J. et al. The genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. PNAS 87: 4707-4711 (1990).

I have "mutation" and "transcription" as keys in search profile. "DNA-damage" is conceptual equivalent of "mutation"; and I know now to link the phrases -- as would be self-evident in a conceptual tree.

3) Bach, R., Friedland, P., and Iwasaki, Y.: Intelligent computational assistance for experiment design. Nucleic Acids Res. 12(1):11-29, January, 1984.

 Karp, P.: Hypothesis Formation by Design, in Computational Models of Scientific Theory Formation, edited by J. Shrager and P. Langley, 1989, in press.