

✓ → DT.
OK CT 1/2/91



MARINE BIOLOGICAL LABORATORY

WOODS HOLE MASSACHUSETTS 02543 (508) 548-3705



(508) 548-1685
email:
Mriley@hoh.mbl.edu

December 21, 1990

Dr. Joshua Lederberg
Rockefeller University
1230 York Ave.
New York, NY 10021

Dear Dr. Lederberg,

I wonder if you remember me. I changed overhead transparencies for you when you gave a talk in Woods Hole a couple of summers ago.

I'm writing to ask for your thoughts on a project I want to undertake -- to start building an electronic database on metabolism. I am interested in whether you see any value in having a metabolic database, and if so, what use you think it might have at early stages of development as well as in a mature form.

I have joined with a couple of computer scientists to set up a "knowledge base", a flexible and sophisticated type of database, using data that I would assemble from the literature. I plan to start with central metabolism, using bibliographic searches and reviews to locate the main journal articles characterizing each enzymatic reaction. I enclose two filled in templates for two enzymes of E. coli, as an example of our early efforts. Most citations are given as bracketed numbers representing ID numbers in Medline. As we go along, accumulating information, I plan to submit the filled-in templates for each reaction to respective experts for validation or correction and addition. Information on the compounds would come from a database being assembled by my collaborator, Peter Karp, containing the CAS Registry #, Mol. Wt., empirical formula and a simple planar structural representation.

We are looking to the new Database Analysis Program at NSF for funding. They tell us that they will not be able to provide support for our entire grand plan at this time, but will consider supporting a small initial project with a well-defined goal and potential usefulness, to serve as a prototype project. Although we see clearly the ultimate goal of establishing a knowledge base handling much of the information on metabolism, we find it more difficult to define valid short term goals for early, relatively simple versions of this database.

We are not certain whether it might be most useful to work toward attaching the assembled metabolic information to emerging genomic

databases, such as a database of E. coli genetic information or the mouse encyclopedia, or whether a metabolic database would be more useful to model alternate pathways, with applications in the food and pharmaceutical industries. In its early, rather simple form, the database might contain little more information than can be acquired easily from a textbook. How could it be useful in early stages? Would it be useful as an educational tool?

Should the database contain information about enzymes from only one organism such as E. coli or mouse or human, or should it contain information on the best characterized enzymes, independent of source?

I think of you as a broad thinker, viewing biology as a whole, a person who would have some reactions to the overall merits of this project, a project that would ultimately put information on the huge field of metabolism into easily accessible, searchable and relatable form. Maybe you can help us focus the aims, particularly the short term aims, of this project. These are the ones I have trouble seeing. I would greatly appreciate any help you can give. Feel free to use phone or email.

Sincerely yours,

Monica

Monica Riley