T-21

From Stepping "Stones to Stumbling" Blocks

How yesterday's intellectual summits become impediments to new creative advances in science.

Joshua Lederberg (RU) and Richard Zare (Stanford) will coedit and contribute to a volume of commentary, exposition and foresight.

We anticipate soliciting chapters from up to a dozen of our colleagues in various branches of science (incl. engineering and medicine) each analyzing their own field. We would also poll 100 or 200 to get their short lists of suggestions. These will be tabulated and annotated by the editors.

Each chapter will recount the history of about ten leading doctrines in science, each of which has been shown to be flawed and now by common consensus abandoned. In most cases, each doctrine was a hardwon advance: e.g., "enzymes are proteins". But it then became institutionalized in a way that hindered further advance -- as we have but recently learned (to win Nobel Prizes for Cech and Altman) some enzymes are RNA. The focus will be on cases that are no longer controversial, and which offer a well authenticated historical record. For the most part we will be avoiding current, unsettled controversies and especially those represented by aggressive, well-organized "schools" as in the social sciences.

In addition each contributor will be asked to suggest a few doctrines not generally regarded as controversial, but which he/she urges be put on the agenda for reexamination.

Appended is a rough draft outline of JL's chapter. This was described in the context of a research program on expert systems in molecular biology. It became evident that expertise should not be taken for granted -- it might be flawed if the precepts are uncritically adopted; we intend an anti-expert system.

T/15

Quis custodiet ipsos custodes? [Juvenal, Satires, vi. 347]
Who watches the watchers?

Joshua Lederberg, University Professor, The Rockefeller University.

(Adapted from an impromptu presentation at the Spring AAAI meetings, AI - molecular biology, Stanford University, 29 March, 1990)

One of the most difficult steps in the development of an expert system is the recruitment and exploitation of the domain wizards. Almost always it is necessary to establish teams of specialists to deal with the programming issues and the user interfaces as well as the incorporation of domain specific knowledge. Experts will communicate how they read a gel, or what is the canonical biological interpretation of DNA sequences conserved over phyletically diverse organisms. The computer scientist will rarely have an independent base of knowledge and experience for critical judgments about the wisdom thus received.

Therein may lie the greatest hazards from the proliferation of expert systems; for much of that expertise is fallible.

It is 12 years since I have been actively involved in the collaborations that led to the DENDRAL and MOLGEN projects (1,2); and I am just now at an early stage of planning a resumption of research on theory formation and validation, as applied to molecular biology. But I recall how easily the most primitive errors could become locked into firm rules -- which would sometimes persist for a long time until revealed by lucky accident. For example, we had what we called a BADLIST in DENDRAL, intended to filter out substructures that experience told were unstable or otherwise untenable. This can give enormous economy in pruning back a combinatorial explosion. One such rule was quite plausible: BADLIST included a proscription against substructures with 2 -NH2 (amino) groups pendant on a single carbon. C..(NH2)2 can be expected to split off ammonia. But one of us overlooked two outstanding exceptions, namely urea and guanidine, (NH2)- C:O - (NH2) and (NH2)-C:NH-(NH2) -- we were so fixated on prohibitions that would apply quite successfully to much larger molecules.

I intend, however, to put that self-skepticism to a larger, constructive purpose. My first target is an examination of many of the central doctrines in the history of micro- and molecular biology, especially those that we have learned to have led us to egregious error. (See Chart 1.) I call those the "Myths we have lived and died by". By and large they are half-truths whose domain of veracity and application was perceived to go far beyond the evidentiary basis that led to their adoption. And we cannot live with prolonged suspension of disbelief in these myths, or we would be practicing nothing but an unremitting nihilism.

I will examine the logical structures that founded the adoption of these beliefs, and again the data and reconstructions that led to their demise. This will require a system of knowledge-representation that will enable a more formal examination of these theories, and in turn a computer based system for critical scrutiny (theorem-proving) and new hypothesis generation.

All of this work is a direct extrapolation of the DENDRAL effort, which used essentially the same approach for "theories" (postulated chemical structures) in the more readily formalizable domain of organic chemical analysis. There the data came originally from mass spectrometry and NMR; later we developed a more flexible interactive system (CONGEN) that enabled all source inputs. One of the interesting uses of CONGEN was as a theorem-prover, namely to reexamine the purported proofs of structure that had been published in a leading journal of organic chemistry. You guessed it, many of those proofs were at least formally defective; and in at least one case that had eluded the human reviewer, substantively so.

Chart 1

The myths by which we live and die.

BACTERIA are SCHIZOMYCETES
i.e., divide only by fission. But Lederberg (1946) showed they had sex

BACTERIA REPRODUCE SEXUALLY (+)
But Lederberg (1951) took that too literally and missed the unique mechanisms of progressive DNA transfer (takes 100 minutes!) discovered by Jacob.

TOXINS KILL
an important paradigm in history of infectious disease. But Koch and the world was misled for 80 years in searching for the "cholera toxin" as an agent lethal by parenteral assay. That toxin "merely" promotes the secretion of water into the gut. The misunderstanding has cost 10s of millions of lives that could have been saved by feeding salt water.

DNA --> RNA
overlooked the reverse transcriptase (DNA <-- RNA), earned a Nobel Prize for Baltimore and Temin.

COLINEARITY OF DNA WITH PROTEIN (1:1 theory)
and ENZYMES are PROTEINS
Classic work of Beadle & Tatum; Benzer; Yanofsky.
Overlooked m-RNA processing, introns. Earned Cech a Nobel prize (for ribozymes)

ONLY GERM CELLS MATE
But somatic cells can be fused too (Lederberg 1955), and enable somatic cell genetic analysis

THE SOMA INHERITS THE GENOME OF THE ZYGOTE FROM THE TWO PARENTS
(Cf Weismann's dogma above). This is already known to be a half truth vis a vis "imprinting" of the paternal contribution. Genetic diversification is central to the mechanisms of antibody formation (Lederberg, 1959). There is increasing evidence for other exceptions, viz. some "epigenetic" changes involve structural alterations in the DNA (Yokota et al. 1989). {N.B. this obviously has large implications for the *definition* of what we mean by the human genome.}

MUTATIONS ARE DELETERIOUS
Circular reasoning: most visible mutations are visible. But 99% of nucleotide substitutions are invisible. Delayed evolutionary theory of drift (Kimura) and engenders gross miscalculations of the genetic disease load attributable to mutation.

GENES HAVE A FIXED LOCUS; SEGREGATE 1:1 (Mendel onward)
But some genes jump! (McClintock)
Segregation is not so rarely perturbed by "gene conversion"

INFINITUDE OF ANTIBODIES - and Pauling's instructionist theories
Slowed up clonal selection theory, now accepted for antibody formation

TETRANUCLEOTIDE DNA - PA Levene's model
was at most a tentative recapitulation of primitive data, but taken too rigidly greatly delayed the recognition of DNA as the genetic material

CHEMICALS CAUSE CANCER
a simplicism that greatly oversimplifies the multifactorial basis of carcinogenesis, and leads to enormous misfocus in managing environmental hazards.

LIFE EVOLVED ON EARTH - (Oparin, Miller-Urey)
but chemical evolution probably started with cosmic condensation.
Open possibility: all organic material on earth is derived from cometary and meteoritic infall, may now be leading hypothesis.

---- With a few exceptions I have been personally involved in these bifurcations. At least once (+) to my chagrin!!

(1) Lindsay, R.K., B.G. Buchanan, E. A. Feigenbaum and J. Lederberg Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project. McGraw-Hill Book Co., (1980).

(2) Yokota H; Iwasaki T; Takahashi M; Oishi M A Tissue-Specific Change in Repetitive DNA in Rats Proceedings of the National Academy of Sciences of the United States of America, 86, (23): 9233-9237 (1989)

Forsythe-DE Buchanan-BG Knowledge Acquisition for Expert Systems - Some Pitfalls and Suggestions IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS Vol 19 Iss 3 pp 435-442 1989 (AM016)

Karp,P. 1989 A process-oriented model of bacterial gene regulation. Stanford Univ. KSL-88-18

Karp,P. and Friedland,P. 1989? A conceptual reconstruction of the discovery of attenuation. SU KSL- ...

N Q335/D263 Davis, Randall; Lenat, Douglas B Knowledge-based systems in artificial intelligence. 490 p. New York: McGraw-Hill. 1982.

CN Q175/L283 Langley, Pat; Simon, Herbert Alexander; Bradshaw, Gary L; Zytkow, Jan M Scientific discovery: Computational explorations of the creative processes. 357 p. Cambridge MA: MIT Press. 1987.

Kulkarni-D Simon-HA ## The Processes of Scientific Discovery - The Strategy of Experimentation COGNITIVE SCIENCE Vol 12 Iss 2 pp 139-175 1988 (P3194)

Stefik-M ## Planning with Constraints MOLGEN .1. ARTIFICIAL INTELLIGENCE Vol 16 Iss 2 pp 111-139 1981 (UR762)

Friedland-P Kedes-LH ## Discovering the Secrets of DNA COMMUNICATIONS OF THE ACM Vol 28 Iss 11 pp 1164-1186 1985 (ATS36)

Friedland-P Kedes-L ## Discovering the Secrets of DNA COMPUTER Vol 18 Iss 11 pp 49-69 1985 (ATZ84)

My intention is to review the principal doctrinal themes of molecular biology from a similar perspective. But armed with an easy retrospectroscope, I thought it only fair to be put on the line for some as yet unsubstantiated future revulsions of thought. These are to illustrate objectives. As yet I have done no explicit programming on this issue. Nevertheless, I have found great value in the style of thinking that is evoked in the context of designing the computer systems. (Harking back to DENDRAL, it also led to a style of critical mental chemistry that matches in importance the first order assistance from the machine.)

So here are three intended bona fides -- Contradictions to the existing regime of thought that, I believe, will be experimentally tested in the near future. Both of them are deeply embedded in the conventional wisdom!

A) *1) The 3-dimensional shape and functionality of (folded) proteins is fully determined by the primary amino-acid sequence, and this in turn by the nucleotide sequence of the gene.* [The latter part of this statement is already eroded by knowledge of messenger RNA splicing, and further by some remarkable examples of post-transcriptional editing of RNA]. This doctrine has been essential for the development of mechanistic ideas of cell and organelle assembly, and especially for our modern views of antibody formation (2).

BUT, this is probably an overstatement. My counter-prediction is that we will discover examples where ambiguous and divergent patterns of folding will enable a given primary protein sequence to fold into two or more well defined, and biologically distinctive final conformations. It is hard for me to imagine that evolution has not exploited this potentiality for flexibility in use of a given blueprint. Evidence for this has been counter-selected, and often discarded as precipitates or "noise". A number of experts of folding have agreed, that

"yes", this should be more carefully considered.

What a neat regulatory system could be hiding under our demand for "purifying" proteins to crystalline homogeneity, thus obscuring allomorphisms. Chaperones might well guide the folding to one or another metastable conformation. (This is not quite the same as allosterism, where a given ligand reversibly alters conformation; allomorphs are kinetically (meta-) stabilized by intra-molecular forces, by analogy to intra-crystal forces with allomorphic phases of crystals).

B) *2) The germ line in multicellular animals is completely segregated from the soma.* This Weismann's doctrine is the foundation of the refutation of lamarckian and lysenkoist ideas, and perhaps for that reason has never been critically examined, except with the crude anatomical methods of the last century. It is certainly very nearly true! However exceptions could be of critical importance, for evolution, pathology, and biotechnology.

C) *3) ENZYMES CATALYZE CHEMICAL REACTIONS., viz. are not consumed..* In fact, experimentally contrived suicide substrates are designed to titrate the corresponding enzymes, and these are perforce "consumed". This is a well known laboratory artefact; what does not come to mind, but I have not searched, is any incorporation of that principle into normal physiology. (That would be an "enzyme" that reacts stoichiometrically and irreversibly with certain natural substrates. The reaction of methemoglobin with cyanide is a near miss.)

I am seeking a still more systematic way to discover issues where a computer-aided custodian could be a help, not of mere incremental advance, but of further scientific and technological revolutions.