Cell, Vol. 50, 667, August 28, 1987, Copyright © 1987 by Cell Press

Letter to the Editor

"Homology" in Proteins and Nucleic Acids: A Terminology Muddle and a Way out of It

"Homology" has the precise meaning in biology of "having a common evolutionary origin," but it also carries the loose meaning of "possessing similarity or being matched." Its rampant use in the loose sense is clogging the literature on protein and nucleic acid sequence comparisons with muddy writing and, in some cases, muddy thinking.

In its precise biological meaning, "homology" is a concept of quality. The word asserts a type of relationship between two or more things. Thus, amino acid or nucleotide sequences are either homologous or they are not. They cannot exhibit a particular "level of homology" or "percent homology." Instead, two sequences possess a certain level of *similarity*. Similarity is thus a quantitative property. Homologous proteins or nucleic acid segments can range from highly similar to not recognizably similar (where similarity has disappeared through divergent evolution).

If using "homology" loosely did not interfere with our thinking about evolutionary relationships, the way in which we use the term would be a rather unimportant semantic issue. The fact is, however, that loose usage in sequence comparison papers often makes it difficult to know the author's intent and can lead to confusion for the reader (and even for the author).

There are three common situations in which hazards arise by using "homology" to mean similarity. The first case is the most obvious offense but perhaps the least troublesome. Here an author identifies sequence similarities (calling them homologies) but claims that the sequences being compared are not evolutionarily related. Some awkward moments occur in such a paper, since the author claims both homology (i.e., similarity) and nonhomology (i.e., lack of a common ancestor). Nonetheless, the author's ideas are likely to be clear since arguments against common ancestry are presented explicitly.

A second case is one in which an author points out similarities (again called homologies) but does not address the issue of evolutionary origins. The reader, seeing the term "homology," may infer that the author is postulating coancestry when that is not the author's intent.

The final case occurs most frequently and is the most subtle and therefore most troublesome. Here, similarities (called homologies) are used to support a hypothesis of evolutionary homology. In this case, the two meanings of homology seem to overlap, and it is almost inevitable that the thinking of author and reader alike will be intrusively distorted as follows. Similarity is relatively straightforward to document. In comparing sequences, a similarity can take the form of a numerical score (% amino acid or nucleotide positional identity, in the simplest approach) or of a probability associated with such a score. In comparisons of three-dimensional structures, a typical numerical description is root-mean-square positional deviation between compared atomic positions. A similarity, then, can become a fully documented, simple fact. On the other hand, a common evolutionary origin must usually remain a hypothesis, supported by a set of arguments that might include sequence or three-dimensional similarity. Not all similarity connotes homology but that can be easily overlooked if similarities are called homologies. Thus, in this third case, we can deceive ourselves into thinking we have proved something substantial (evolutionary homology) when, in actuality, we have merely established a simple fact (a similarity, mislabeled as homology). Homology among similar structures is a hypothesis that may be correct or mistaken, but a similarity itself is a fact, however it is interpreted.

We believe that the concepts of evolutionary homology and sequence or three-dimensional similarity can be kept distinct only if they are referred to with different words. We therefore offer the following recommendations:

- Sequence similarities (or other types of similarity) should simply be called similarities. They should be documented by appropriate statistical analysis. In writing about sequence similarities the following sorts of terms might be used: a level or degree of similarity; an alignment with optimized similarity; the % positional identity in an alignment; the probability associated with an alignment.
- Homology should mean "possessing a common evolutionary origin" and in the vast majority of reports should have no other meaning. Evidence for evolutionary homology should be explicitly laid out, making it clear that the proposed relationship is based on the level of observed similarity, the statistical significance of the similarity, and possibly other lines of reasoning.

One could argue that the meaning of the term "homology" is itself evolving. But if that evolution is toward vagueness and if it results in making our scientific discourse unclear, surely we should intervene. With a collective decision to mend our ways, proper usage would soon become fashionable and therefore easy. We believe that we and our scientific heirs would benefit significantly.

Gerald R. Reeck,[1] Christoph de Haën,[2] David C. Teller,[3] Russell F. Doolittle,[4] Walter M. Fitch,[5] Richard E. Dickerson,[6] Pierre Chambon,[7] Andrew D. McLachlan,[8] Emanuel Margoliash,[9] Thomas H. Jukes,[10] and Emile Zuckerkandl[11]

[1]Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506; [2]Departments of Medicine and Biochemistry, University of Washington, Seattle, Washington 98195; [3]Department of Biochemistry, University of Washington, Seattle, Washington 98195; [4]Department of Chemistry, University of California at San Diego, La Jolla, California 92093; [5]Department of Molecular Biology, University of Southern California, Los Angeles, California 90089; [6]Molecular Biology Institute, University of California, Los Angeles, California 90024; [7]Institute of Biological Chemistry, Strasbourg, France; [8]MRC Laboratory of Molecular Biology, Cambridge, England; [9]Department of Biochemistry and Molecular Biology, Northwestern University, Evanston, Illinois 60201; [10]Department of Biophysics, University of California, Berkeley, California 94720; [11]Linus Pauling Institute, Palo Alto, California 94306