nature biotechnology

Please send letters to:
Nature Biotechnology
345 Park Avenue South
New York, NY 10010-1707, USA
Fax: +1 212 696 9635
E-mail: biotech@natureny.com

The hydropathic binary code: A tool in genomic research?

To the editor:

An editorial in the June issue addressed the implications of the new genomic company planned by Perkins-Elmer and Craig Venter which will complete sequencing of the human genome within three years. This acceleration of the sequencing job will increase the pressure for academic and commercial institutions to develop and use informatic tools for the mining of genomic sequence data which, in turn, will become a critical factor for the survival of genomic biotechnology companies.

In the design of computer-based mining tools, no attention has been paid to a unique feature in the genetic code that determines the basic physicochemical character of the encoded amino acids. Blalock et al.[1], recognizing the importance of inverted hydropathy in protein-protein interaction, analyzed the underlying genetic basis of inverted hydropathy and proposed in 1985 a molecular recognition theory which postulates that complementary nucleic acid sequences encode peptides or proteins which interact. According to this theory the interaction is mediated by amino acid sequence structures with complementary hydropathic profiles.

The underlying basis of the hydropathic complementarity of interacting proteins encoded by sense and antisense DNA strands is a unique feature of the genetic code that has escaped much attention. A binary code divides amino acids into two groups, one with positive hydropathic scores and the other with negative scores, as determined by Kyte and Doolittle[2]. A purine base in the second position encodes amino acids with hydrophilic character, while pyrimidine bases code for hydrophobic or hydropathically neutral amino acids. Hence, DNA sequence regions rich in complementing nucleotides in the second triplet position code for protein regions with inverted hydrophathies.

The hydropathic binary code has been used extensively to design and predict interacting proteins or peptides[3]. A model for interacting complementary structures postulates secondary structures, β-strands or α-helices, in which the hydrophilic surfaces are oriented towards the aqueous phases, while the hydrophobic surfaces are facing each other[4].

Recently it occurred to us[5] that the binary code, which controls the hydropathy of amino acids and thereby predicts where proteins bind to each other and may functionally interact, could be used to scan DNA sequence data, including complete gene sequences and DNA fragments cloned with expression sequence tags (EST). The feasibility of such searches has recently been demonstrated. DNA sequence regions encoding for protein regions with complementary secondary structures have been found and identified within a single gene[6]. These DNA segments have been called "antisense homology boxes" (AHBs). If AHBs with the potential of reciprocal binding are to be derived from two different interacting proteins, they should be surface exposed on each protein. Thus, a search for DNA sequences coding for proteins with inverse hydropathies[2] would predict that these sequences are from genes coding for proteins with the potential to make contact and interact.

In addition, this method to scan the DNA sequence banks for regions with identical or similar binary codes identifies evolutionary conserved sequences, and thus would provide a new tool in "comparative genomics."

Heinz Kohler
University of Kentucky
Lexington, KY 40536-0096
(hkohl00@pop.uky.edu)

Edwin Blalock
University of Alabama
Birmingham, AL 35294-0005
(blalock@phybio.bhs.uab.edu)

1. Bost, K.L. et al. 1985. *Proc. Natl. Acad. Sci. USA* **82**:372–375.
2. Kyte, J.A. and Doolittle, R.F. 1982. *J. Mol. Biol.* **157**:105–132.
3. Blalock, J.E. 1990. *Trends Biotech.* **9**:140–144.
4. Kamtekar, S. et al. 1993. *Science* **262**:1680–1685.
5. Kohler, H. and Blalock J.E. U.S. Patent #552308.
6. Baranyi, L. et al. 1995. *Nature Med.* **1**:894–901.

IGF in the clinic

To the editor:

In response to John Fowlkes' letter concerning my "one-sided analysis" of IGF-1 in the March issue, I would propose that in fact it is his analysis which is biased in favor of the drug. In my article, I presented the research of numerous scientists who had used IGF-1 in a variety of clinical settings, all of which has raised serious questions about whether its risk/benefit ratio is sufficiently favorable to warrant clinical use.

While, as Fowlkes asserts, no drug is without side-effects, and granted, IGF-1 may have utility in treating type I diabetes, one must ask whether it is as safe and effective for that indication as other drugs. The FDA raised questions concerning Genentech's trials in these indications and long-term safety, which we believe were conducted at Fowlkes' university. If Genentech believed in the clinical efficacy and safety of the drug—and stood behind it without reservations—one must ask why then did it choose to suspend its advanced-stage trials following the FDA's request for long-term safety studies?

No one disputes that IGF-1 may be somewhat effective in treating type I diabetes, growth hormone deficiency, or other diseases, but is it the best treatment with the least amount of risk? Is it as good as other treatments? No research has to date shown either to be true. While Fowlkes asserts that type I diabetics showed "marked improvement in glycemic control," and "when used in IGF-1-deficient populations, [it has] beneficial effects," no data exists that shows it to be safe with long-term use, and with Genentech withdrawing it from trials, chances are great that its putative long-term safety will not be shown. Further studies in these areas are not taking place because Genentech has chosen not to conduct them. We should not blame the FDA or other researchers' findings on IGF-1 if there is a lack of long-term safety data on the drug.

Although Pollack's study was not a trial using IGF-1, but rather examined the relationship between plasma levels of IGF-1 and prostate cancer risk, to say that his study shows no cause-and-effect relationship between the two would be akin to saying that a study showing a relationship between elevated cholesterol levels and heart attack risk is invalid because cholesterol was not given as a drug in such a study. Plasma levels of many substances are commonly measured to uncover disease and risk factors for disease; measuring plasma levels is not "com-

Editorial clarification

In an editorial in the June issue of *Nature Biotechnology* (16:491) "Keeping the story straight," sponsorship of a bill allowing for the development of an integrated database for gene discovery research on the Icelandic population ("bill on medical databases") was attributed to Kari Stephansson, CEO of deCODE Genetics. While Dr. Stephansson was instrumental in developing the concept of the integrated database that is part of this legislation, the bill itself was formally sponsored by and put before the Icelandic parliament at the end of March by the minister of health on behalf of the Icelandic government. ///