CORRESPONDENCE

From syntax to semantics

To the editor:
Your editorial "Collective unconsciousness" (*Nature Biotechnology* 15:481, June 1997) raises the point of the increasing diversity in the scope and activities of biotechnology companies. This is even more striking and interesting when one includes in the description of the "biotechnology sector" companies and departments dedicated to the new information-intensive era of biomedical research and drug development. Biotechnology, in that sense, is quickly becoming an equivalent of the high-technology and digital media sector, with its content providers (e.g., sequences, expression profiles), infrastructure builders (e.g., bioinformatics databases and networking), and enabling technologies (e.g., gene chips, high-throughput robotics environments, software for structural analysis, and prediction).

The current wave of bioinformatics companies and departments within larger companies is only the starting point of an upcoming, far more significant collaboration between biotechnology and drug development on the one hand, and information and computer sciences on the other. Today's content is quantitatively impressive but qualitatively and by nature of low level. Be it to build or access databases, or to search the genome, it is mainly focused on the syntactic level of information. The use of existing or emerging bioinformatics infrastructures is further constrained by the inherent brittleness of databases due to incomplete information and inconsistent annotations. Still, enabling technologies like gene chips are about to pour out another flood of information, which will have to be analyzed from the perspectives of many different and yet unrelated fields.

The real challenge to which life science computing has to scale up is the passage from syntax to semantics. As one reaches beyond sequencing and the now-traditional algorithms that provide some early form of interpretation to functional genomics, discovering the real context of expression, function, and integration in a whole that is much larger than its parts will require new approaches. Serious hints already abound that the genomics approach is running into some real but predictable difficulties. Homology and function are not trivially related, and biochemistry and biology are becoming essential

in interpreting results. Expression profiles do not necessarily correlate with disease. Identification of a gene does not mean a direct path to diagnostic tests, as most recently seen with the *BRCA1* gene and breast cancer[1]. Contrary to some expectations, loads of information on the genome does not magically simplify the process of drug discovery and development, but rather, makes it fundamentally different. The heuristics that will reduce the complexity of searching for answers lie not just in low-level information analysis but also, and probably more so, in the richness of the interactions between all the disciplines involved, such as biology, pharmacology, or clinical medicine. The process of discovery will shift from one in which disciplines are sequentially involved to one in which they are working in parallel and interacting.

For biocomputing scientists, this "info age biotech," means reaching beyond the better-known settings of sequence analysis and annotation to address the semantics of the full genes-to-diseases (and back) spectrum. Informatics should play a key role in helping to make sense of the complex contingencies involved. It should, for example, enable researchers to formulate new and creative experimental strategies and help physicians to propose appropriate patient stratifications and design preventative and therapeutic approaches based on genetic variations, polymorphisms and epigenetic factors. In the resulting new value chain, content will extend from sequences and query results to working models, theories, and other qualitative analyses to assist discovery efforts.

There can be no doubt that the scope and diversity of the "biotechnology sector" is about to increase even further, as it becomes more information-intensive. Not only should this diversity be acknowledged, but as information and computing solutions start to address the semantics of biomedical information, it will become instrumental in shifting the current paradigm of drug discovery to a multidisciplinary one in which the triad of prevention, diagnostic, and therapeutics is more fully realized for more diseases and their variations than ever before.

Alain Rappaport
Carnegie Mellon University
School of Computer Science
Pittsburgh, PA 15213
(atr@cs.cmu.edu)

1. Couch, F.J. et al. 1997. *New Engl. J. Med.* **336:**1409–1416.

Debunking hCG

To the editor:
We were intrigued by the news report in *Nature Biotechnology* (15:834–835, September 1997) that researchers including Robert Gallo were now starting to recognize that they

had been incorrect in thinking human chorionic gonadotrophin (hCG) has anti-HIV and anti-Kaposi's sarcoma activity. We had reached the same conclusion some time ago.

In the 1990s, claims emerged that hCG had anti-HIV[1] and anti-Kaposi's sarcoma action[2]. The reports from Gallo's National Institutes of Health group suggesting that hCG has curative effects on AIDS-related Kaposi's sarcoma claims were based on several misconceptions and inaccuracies.

Using massive amounts of impure hCG preparations for local injections, the NIH group reported reduction in tumor size in AIDS-KS patients, raising expectations of finding another use of hCG[3]. They have now recognized that these effects were due not to hCG or to its β subunit as they had claimed. They attribute them now to a partially characterized substance they call HAF (hCG-associated factor).

However, we had immediately realized, on examining the 1995 report of the Gallo group[2], that the effects might not necessarily be due to hCG or its subunits. The clinical grade material they used (which is extracted from human pregnancy urine) is only about 25% pure and any effects could not be attributed to the hormone unless supported by other data. Working independently and using the same AIDS-KS cell line KSY-1 as the NIH group, we knew by March 1996 that pure hCG had no effect on tumor cells[4]. Our next step, based on previous work[5] was to fractionate crude hCG to locate the active fraction.

In the September issue of *Endocrinology*[6], we documented that the anti-KS agent in pregnancy extracts is neither hCG, nor its α or β subunits. None of four highly purified hCG preparations, including a recombinant hormone, showed any inhibitory effect on the growth of KSY-1 cells in vitro. The anti-KS action is not associated with a macromolecular entity unlike hCG or its subunits but emerges in a fraction associated with small molecules. We found evidence that the anti-KS activity in crude hCG could be acting by downregulating a transcription factor complex called AP-1. This suggests a genomic target. Our report resolves a controversy and suggests that an unidentified molecule(s) may be the active substance.

M. Ram Sairam[1] and Tony Antakly[2]
[1]*Molecular Reproduction Research Laboratory*
Clinical Research Institute of Montreal
& [2]*Department of Biochemistry*
University of Montreal, Quebec, Canada
(sairamm@ircm.umontreal.ca)

1. Bourinbaiar, A.S. et al. 1992. *FEBS Lett.* **309:**82–84.
2. Lunardi-Iskander, Y. et al. 1995. *Nature* **375:**64–68.
3. Gill, P. et al. 1996. *New Engl. J. Med.* **335:**1261–1269.
4. Antakly, T. et al. 1996. *Proc. Am. Assoc. Cancer Res.* **37:**Abst. #1543.
5. Manjunath, P. and Sairam, M.R. 1982. *J. Biol. Chem.* **257:**7109–7115.
6. Kachra, Z. et al. 1997. *Endocrinology* **138:**4038–4041.