LETTERS

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted by e-mail (science_letters@aaas.org), the Web (www.letter2science.org), or regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Planning the Genome Institute's Future

ELIZABETH PENNISI'S ARTICLE "GENOME institute wrestles mightily with its future" (News Focus, 29 Nov., p. 1694), on the National Human Genome Research Institute's recent meeting to unveil its draft plan for its future, missed the mark in capturing the accomplishments and tenor of the meeting. Approximately 200 participants with diverse areas of expertise were invited to critically assess and provide input on all aspects of NHGRI's ambitious, comprehensive, and far-reaching draft plan. The dissension over the plan had less to do with the content and more to do with the format of the presentation. The attendees were excited about the vision of the future that was described but argued that there needed to be a clearer articulation of the priorities, timeline, and NHGRI's degree of involvement. The draft plan centered around three pillars: (i) Genomics to Biology: Elucidating the Structure and Function of Genomes, (ii) Genomics to Health: Translating Genome-Based Knowledge into Health Benefits, and (iii) Genomics to Society: Promoting the Use of Genomics to Benefit Society. There was consensus that the representation of the plan as pillars did not work—pillars were viewed as overly rigid and noninteracting. During the meeting, this representation of the plan was successfully reworked. The pillars became three floors of a house, resting on a foundation of the Human Genome Project, with "Genomics to Society" on top. Key components then stretched across multiple floors: "ELSI and Policy," "Education," "Workforce," "Technology and Methods," "Information Science," and "Resources." NHGRI is to be commended for seeking input from a diverse group of stakeholders. It is only through such discussions and thoughtful strategic planning that society will realize the scientific and public health benefits of the Human Genome Project.

WENDY R. UHLMANN,[1] ROBIN BENNETT,[2] JEFFREY R. BOTKIN,[3] DAVID BOTSTEIN,[4] JOANN A. BOUGHMAN,[5] ARAVINDA CHAKRAVARTI,[6] ELLEN WRIGHT CLAYTON,[7] JEFFREY KAHN,[8] BARBARA KOENIG,[9] THOMAS H. MURRAY,[10] MAYNARD V. OLSON,[11] JANET ROWLEY,[12] SHARON TERRY,[13] DAVID VALLE[14]
[1]Division of Molecular Medicine and Genetics, University of Michigan, 4301 MSRB III, Ann Arbor, MI 48109–0638, USA. E-mail: wuhlmann@umich.edu. [2]National Society of Genetic Counselors, Wallingford, PA 19086–6617, USA. [3]University of Utah, Salt Lake City, UT 84113, USA. [4]Stanford University School of Medicine, Stanford, CA 94305–5120, USA. [5]9650 Rockville Pike, Bethesda, MD 20814, USA. [6]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA. [7]Vanderbilt University, Nashville, TN 37232–0165, USA. [8]Center for Bioethics, University of Minnesota, Minneapolis, MN 55455–0346, USA. [9]Stanford Center for Biomedical Ethics, Palo Alto, CA 94304, USA. [10]The Hastings Center, Garrison, NY 10524–5555, USA. [11]University of Washington, Seattle, WA 98195, USA. [12]University of Chicago, Chicago, IL 60637, USA. [13]Genetic Alliance, Washington, DC 02067, USA. [14]Howard Hughes Medical Institute and The Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

Response

THE NATIONAL HUMAN GENOME Research Institute (NHGRI) wisely develops long-term plans with input from researchers from many walks of science. It has been a genome community tradition since before there was a Human Genome Project, and NHGRI should indeed be commended for it.

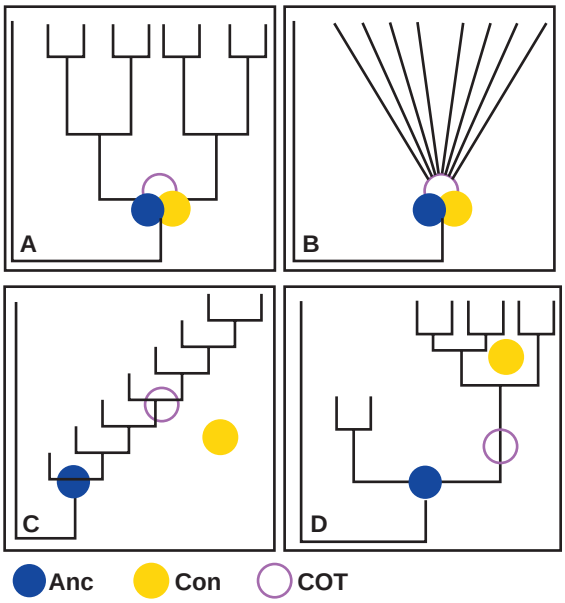
The format of the NHGRI draft presented to the November 2002 meeting was very confusing, however. Conference organizers made clarifying it a top priority. After working almost all night, they came up with a new format—the metaphor of floors in a building that Ullmann and her colleagues describe—that was "inspirational" in that it better communicates what NHGRI and its constituents propose to do.

What were not fixed, and not really fully addressed by the group, were the issues covered in my article, such as how far into clinical research NHGRI should venture, or if the institute can really take on all that it proposes to do in the next 5 years. Those questions were raised by many participants both in public and private. Their resolution, not how the plan will look on paper, is what will determine NHGRI's effectiveness in the post–human genome era.

ELIZABETH PENNISI

Consensus and Ancestral State HIV Vaccines

THE REVIEW BY B. GASCHEN *ET AL.* ("DIVERSITY considerations in HIV-1 vaccine selection," 28 June, p. 2354) describes two computational methods (consensus and ancestral state) being considered for developing vaccine antigens against HIV. These methods attempt to minimize the amount of sequence divergence (distance) between the antigen and contemporaneously circulating viruses. Both methods do well at minimizing these distances when the sequences used to estimate the potential vaccine come from a symmetric phylogeny (panels A and B of the figure), similar to those examined in their Review. However, if sequences used to estimate the potential vaccine come from asymmetric phylogenies (panels C and D), then both methods generate sequences that poorly minimize these distances, making for a potentially poor vaccine antigen.



Four possible phylogenetic shapes and the resulting reconstructed sequences. The Ancestor (Anc) and the Center Of the Tree (COT) can only fall on an evolutionary path (i.e., on a branch of the phylogeny), whereas the Consensus (Con) may not.

We have proposed and championed the ancestral state method (*1*–*3*), believing that it has an advantage over a consensus vaccine because the ancestral state is an estimate of an actual sequence that existed in the past (i.e., it comes directly from the reconstructed history), whereas the consensus sequence need not be, and most likely is not, such an

entity. It is likely that the protein generated from an ancestral sequence will have native folding and function (*4*, *5*), and the ancestral sequence is unlikely to change much as new sequences are added to the sample.

We have developed a third, complementary means of developing a vaccine antigen that has all of the desirable properties of the ancestor but is less sensitive to phylogenetic outliers (e.g., Fig. 1 of Gaschen *et al.*, seq. ZA). This new method identifies a point called point C, Center of the Tree (COT), on the unrooted phylogeny, where the average evolutionary distance from C to each tip on the phylogeny is minimized. Because the COT is a point on the phylogeny, the estimated COT sequence will have the same advantages as the estimated ancestral sequence. In the table, we show that, using the subtype C data presented in Gaschen *et al.*, the COT is at least as similar to sampled amino acid sequences as the ancestor or the consensus, in the *gag* and *env* genes. Although the COT method is as good as either of the other methods at minimizing the differences in the context of subtype C, a real benefit from its use may accrue under different and equally important evolutionary scenarios. By design, the COT will be more similar than the ancestor to viral lineages or clades that have evolved more rapidly than other lineages, perhaps through strong diversifying selection, while maintaining similarity to more slowly evolving lineages. Thus, it may have a particular advantage when phylogenies are found to be more asymmetric, as in the case of trees involving multiple subtypes or recombinant forms.

Artificially generated sequences

Protein	C-Center of the Tree	C-consensus	C-ancestral
p24	96.0 (97.6–90.5)	94.3 (95.6–91.2)	95.3 (98.0–92.2)
gp160	92.1 (94.8–82.0)	92.4 (95.0–82.0)	90.9 (92.7–83.4)

Percentage similarities for the three candidate vaccine sequences, corresponding to Table 2 in Gaschen *et al.* COT, consensus, and ancestral sequences were estimated for HIV-1 subtype C using the same phylogenies and sequences as those in Gaschen *et al.* These artificially generated sequences were then compared with a set of protein sequences from contemporary subtype C isolates and the mean (and range) of divergences reported. Note that differences between these values and those reported in Gaschen *et al.* resulted from different gap stripping criteria.

An effective AIDS vaccine will need to provide broad protection against multiple HIV strains, particularly rapidly evolving ones. Minimizing the number of antigens, while maximizing the breadth of their reactivity, increases the feasibility of development and successful implementation of such a vaccine. To this extent, the COT may address the resulting trade-offs more equitably than either the consensus or the ancestor. Designing a vaccine capable of dealing with the extensive diversity of circulating HIV-1 strains worldwide is a daunting task, particularly in light of recent reports of intersubtype HIV superinfection (*6*). We believe the COT method provides an additional approach to vaccine design that could bridge the gap between highly specific monovalent and multivalent HIV vaccines.

DAVID C. NICKLE,[1]* MARK A. JENSEN,[1] GEOFFREY S.

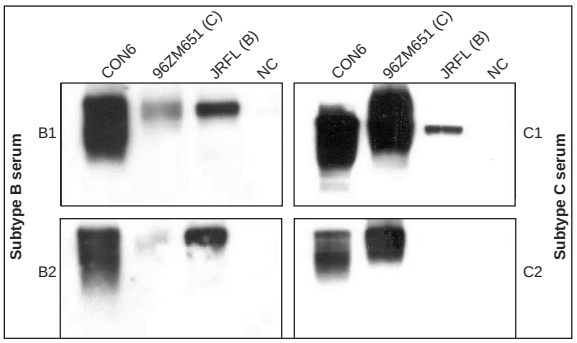
GOTTLIEB,[1,3] DANIEL SHRINER,[1] GERALD H. LEARN,[1] ALLEN G. RODRIGO,[1]† JAMES I. MULLINS[1,2,3]
Departments of [1]Microbiology, [2]Laboratory Medicine, and [3]Medicine, University of Washington School of Medicine, Seattle, WA 98195–8070 USA. *To whom correspondence should be addressed. E-mail: dnickle@u.washington.edu
†Present address: School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand.

References and Notes
1. G. Learn, J. I. Mullins, abstract presented at Twelfth Joint Scientific Meeting of the AIDS Panels for the U.S.-Japan Cooperative Medical Science Program, Santa Fe, NM, 2000.
2. G. Learn, J. I. Mullins, abstract presented at Seventh International Discussion Meeting on HIV Dynamics & Evolution, Seattle, WA, 2000.
3. J. I. Mullins, A. G. Rodrigo, G. H. Learn, N. Doria-Rose, N. Haigwood, paper presented at the HIV-1 Protection and Control by Vaccination meeting, Keystone, CO, 2002.
4. B. S. Chang, M. J. Donoghue, *Trends Ecol. Evol.* **15**, 109 (2000).
5. T. M. Jermann, J. G. Opitz, J. Stackhouse, S. A. Benner, *Nature* **374**, 57 (1995).
6. A. Ramos *et al., J. Virol.* **76**, 7444 (2002).
7. We thank B. Korber and B. Gaschen for very kindly providing sequence data. This work was supported by grants from the United States Public Health Service, including support from the University of Washington Center for AIDS Research.

Response

THE COT METHOD, LIKE THE CONSENSUS AND ancestral strategies (*1*), produces a central sequence suitable for testing as a vaccine candidate. All of these approaches reduce the average distance between a vaccine antigen and circulating HIV-1 proteins. For HIV-1 subtype-specific vaccine strategies, the three derived sequences will be only slightly different, each subject to different biases. The biological properties of vaccines derived from such artificial sequences cannot be predicted. Experiments directly comparing artificial and natural vaccines are needed to test their immunogenicity and extent of cross-reactive protection.

Within-clade HIV-1 sequences tend to give very short internal branch lengths near the basal node of the clade and long branch lengths from the tips to the interior nodes of the tree. In an idealized version of such "star" phylogenies, as illustrated in panel B of Nickle *et al.*'s figure, the ancestor and the COT coincide. In reality, there is some structure near the base of HIV-1 subtype trees (*1*, *2*), resulting in slight differences between the within-clade consensus, COT, and ancestral sequences. The COT and consensus sequences favor heavily sampled sublineages and deemphasize outliers. But in cases where the evolution of the virus is not time-symmetric—e.g., the accumulation of escape mutations in CTL epitopes presented by common HLA types in a given population—



Broad cross-clade reactivity of the M group consensus gp120 with patient sera of different clades. The same amount of M group consensus (CON6), clade B (JRFL), and clade C (96ZM651) gp120 proteins were denatured and bound with HIV-1 subtype B or C sera in a Western blot assay. Four representative examples of eight reactions are shown. Nondenatured gp120 proteins were also tested using ELISA (5), and seven subtype B or C sera bound the M group consensus comparably to within-clade proteins (*4*).

the consensus may produce a sequence bearing epitopes more relevant to the current epidemic than either the COT or ancestral state.

Nickle *et al.* write that "the ancestral state method… has an advantage over a consensus vaccine because the ancestral state is an estimate of an actual sequence that existed in the past." We reemphasize (*1*) that one should be cautious about supposing that reconstructed sequences are meaningfully similar to viral genomes that once existed. Ancestral and COT sequences are both artificial, defined as the most likely state, at a particular interior point in the tree, consistent with the observed data given the evolutionary model. Discordance between biological reality and the evolutionary model can lead to faulty inference of ancestral states. Evolutionary models generally capture the "average" evolutionary pattern over a region under consideration, rather than the nuances of individual sites and lineages, and models generally ignore recombination (*3*), correlation between sites, differences in mutational patterns induced by neighboring bases, and differing selective pressures in different regions and lineages. Such problems are apt to be most severe at sites with the greatest immunological significance, because these sites will be evolving under positive selection and are subject to between-site correlations, whereas the models assume neutral evolution and independence of sites. Also, the most likely ancestral state is, in fact, not very likely. For example, even ignoring errors in the evolutionary model and the tree, the probability that there are no errors in the C-subtype ancestor envelope sequence in our reconstruction (*1*) is 5×10^{-15}, and the expected number of errors is 28 of 2379 sites (1.2%). The difference between the ancestor, consensus, and COT is of the same order as the expected number of errors (1.2%), indicating that there may be little statistical weight to the differences between the three sequences.

The bottom line for incorporating any sort of central, derived protein sequence into vaccines is experimentally determined preservation of key antigenic domains, cross reactivity, and immunogenicity. T cell epitopes are short processed peptides, and so the question of natural versus artificial proteins is of less concern in that context. Antibody epitopes, however, are often conformationally sensitive. We have recently found that a consensus M group gp120 protein bound to soluble CD4 as well as to neutralizing (b12 and 2G12) and conformation-sensitive (17b and A32) monoclonal antibodies (*4*). Furthermore, this M consensus protein exhibited enhanced interclade antigenic cross reactivity (see figure). These results indicate that it is premature to exclude any of the possible central immunogen approaches without rigorous and comparative testing.

F. GAO,[1] T. BHATTACHARYA,[2] B. GASCHEN,[2] J. TAYLOR,[2] J. P. MOORE,[3] V. NOVITSKY,[4] K. YUSIM,[2] D. LANG,[2] B. FOLEY,[2] S. BEDDOWS,[3] M. ALAM,[1] B. HAYNES,[1] B. H. HAHN,[5] B. KORBER[2,6]

[1]Duke University Medical Center, Durham, NC 27710, USA. [2]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [3]Department of Microbiology and Immunology, Cornell University, New York, NY. [4]Department of Immunology and Infectious Disease, Harvard School of Public Health, Boston, MA 02115, USA. [5]University of Alabama at Birmingham, Birmingham, AL 35294, USA. [6]Santa Fe Institute, Santa Fe, NM 87501, USA.

References

1. B. Korber *et al.*, in *Retroviral Immunology*, G. Pantaleo, B. D. Walker, Eds. (Humana Press, Totowa, NJ, 2001), pp. 1–31.
2. V. Novitsky *et al.*, *J. Virol.* **76**, 5435 (2002).
3. A. Jung *et al.*, *Nature* **418**,144 (2002).
4. F. Gao, J. P. Moore, data not shown.
5. J. P. Moore *et al.*, *J. Virol.* **68**, 8350 (1994).