Memorandum

Date 12 April 1989

From Gerald Myers

Subject Questions about sequences

To Dr. Anthony Fauci
Director, NIAID

Early on in the database and analysis project, two problems arose that were unavoidably entangled with the dispute about the discovery of HIV: 1) the database had to select a "reference" sequence that was not necessarily a "prototype" but was at least a paradigm; 2) cluster analyses of variant sequences, an appropriate line of inquiry for the project, immediately drew attention to the close similarity of the IIIb and LAV sequences.

The first problem was for a brief time easy to address because the IIIb clone known as HXB2 was the only infectious molecular clone for which a complete genomic sequence had been determined. Throughout this period, however, the database publication contained disclaimers pertaining to the "primacy" of this virus; we did not want to convey an impression that this virus was (or was not) in the mainstream of the viral population. Of late, it has been more difficult to assume this formal standpoint -- the sample of U.S. isolates has greatly increased, infectious clones are available for many of these, and growing evidence suggests that the IIIb viruses, which have been immensely valuable up to now, may not be so representative or "prototypical" in the future. Thus, beginning with the 1989 compendium of the database, HXB2 will no longer take a prominent role as reference sequence (see appendix 1). Consensus sequences (abstractions) become the closest thing to references.

Regarding the second problem above, the first "tree" analyses were generated at precisely the time (spring 1987) that the U.S. and French were settling the legal dispute that had arisen. Dr. Le Montagne, Dr. Nutter, Dr. Martin, Dr. Rabson and I met and fairly quickly agreed that the database would steer absolutely clear of the issue but that we would not suppress scientific data. Independently, Dr. Wong-Staal, a database editor, and I reached the same conclusion, knowing fully that the "trees" could create new discussion. It was quite a relief to all of us that they did not (for example, see the New Scientist article enclosed as appendix 2), although I remained deeply disturbed about the claim made for the IIIb viruses -- that they derived from pooled blood of several patients. It was very difficult in 1987 to convince many researchers that the AIDS viruses mutated inordinately rapidly. The IIIb interpretation gave the false impression that the virus was more stable than other signs were indicating. (HTLV-1 sequences from remote sources were showing less genetic variability than the IIIb clones, suggesting that the latter were indeed isolates from

several different patients.)

Two years later, the variability question is still thorny, but the matter rests now, as it properly should, in the hands of experimentalists; given what is now known (to a large extent due to the efforts of the NCI researchers), there is no danger, in my opinion, that the variation is being underestimated. However, new problems and questions have emerged, largely resulting from the emerging data obtained by PCR-assisted sequencing:

1) PCR data is not yet statistically robust; moreover, since amplification is the key to the technique, contamination of materials is at least as likely as has been the case with conventional cloning and sequencing.

2) Criteria have not been established for "same and different" viruses, isolates, whatever ("quasi-species" is now a popular word). At stake is the judgment whether a patient has been infected with two viruses from a pool of relatively stable variants or one highly unstable virus. Is superinfection possible?...are "jackpots", akin to what Luria and Delbruck encountered with phage, becoming manifest?

3) Can sequence data be correlated with neutralization studies to determine the major and minor forms of HIVs in a population? (Again, can we assist others as well as guide ourselves with regard to mainstream viruses and minor forms?)

4)Finally, there is the practical problem of collating and publishing sequence data. It is no longer feasible for our project to present each individual sequence. GenBank is committed to doing precisely that and we will be working in close touch with them through the coming months when "buckets" of short sequences will become publicly available.

We are wrestling with these interesting and vital problems --- happily. But these genuine issues focus upon closely related sequences, hence they inevitably tread upon the territory of the 1987 dispute. For instance, we have taken the step to restructure the database to facilitate the new PCR sequence data, and just as importantly, to help establish criteria for relatedness (see appendix 3), and that mere logistical step has immense rhetorical implications. Dr. Wong-Staal and Dr. Josephs, editors from NCI, have been involved in these deliberations and have communicated them to Dr. Gallo and the group at NCI. Dr. Reitz and I have had several valuable conversations about these matters. They have shown nothing but good will and cooperation at every turn. Nevertheless, it is sticky business.

Additionally, Dr. Temin, on behalf of the genetic variation committee, has asked me to undertake analyses necessary for the pursuit of the sampling and sequencing effort; I'll be completing these in the coming weeks, and if you would like to have a copy, do let me know. The ironic twist to this is that only the IIIb and BRU (LAV-1) data is of sufficient robustness to serve as a template for these analyses. In an intitial pass at the problem last fall, it did not escape Dr. Temin's attention that sensitive ground

was being traversed, albeit with ingenuous intentions (see appendix 4). When the final history of early AIDS research gets written, it should be noted that the IIIb/LAV-1 viruses not only served well for the ELISA but also for crucial sequence analyses.

When the variation committee gets in full swing, these problems will be taken up in an orderly fashion. Meanwhile, it is, I think, prudent for you to be appriaed of these areas of concern... and promise. I am more optimistic today about the variation than I have been for two years.

Please let me know if you have any questions.

GM