NEWS ANALYSIS

USPTO: A MISSTEP IN THE RIGHT DIRECTION

ST. LOUIS, Mo.—The U.S. Patent and Trademark Office (PTO) has proposed new rules that most biotech patent applicants submit sequence data on computer disks—following a rigidly defined format.

The proposed rules (37 CFR 1.821-825) appeared in the May 16 Official Gazette and the May 2 Federal Register. If adopted, they would govern applications containing any DNA sequence longer than 10 bases, or any

"To achieve the PTO format, most inventors will have to revise their data."

peptide longer than four amino acids. The rules would affect any patent application that "discloses" or "contains" any such sequence, regardless of whether the sequence is claimed or crucial to the invention (e.g., if you show how a new process or reagent works, using a sequence from a randomly selected gene or protein, the proposed rules would apply). The sequence data must be set off in a separate section, and the printed version would have to be duplicated exactly (as confirmed by a statement made under penalty of perjury) on a 5.25- or 3.5-inch disk in IBM or Macintosh format.

As stated in the PTO notice, "The format proposed herein is based on the Genbank data format and forms currently in use. Submission of sequence data using the current Genbank format and forms is generally acceptable..." Right now, however, Genbank requires only ASCII (American Standard Code for Information Interchange) sequences. It specifies no particular format.

By contrast, the PTO's format specs fill four pages of fine print. For example, non-coding nucleotides must be grouped into segments (separated by spaces) 10 bases long, with no more than 60 bases to a line. The number of the last nucleotide must be inserted into each line and printed in the right margin. Amino acids must be printed on the line above (not

below) the corresponding codons, and every fifth amino acid must be numbered on the line above that amino acid.

To achieve the PTO format, most inventors will have to revise their data, inserting numbers and splicing together DNA and amino-acid sequences exactly as the PTO specifies. The paper version must be printed in a fixed-width font, where the letter "I" takes up the same width as the "M." Typeset fonts (such as the one the PTO used in its example) will be prohibited. On the disk, "hard returns" will be prohibited (word processors vary widely in their treatment of these characters: some use hard returns only to mark the ends of paragraphs; others insert a hard return at the end of every line). That will make setting up patent sequence data using word-processing software very iffy.

If an application violates even a single provision, it can be delayed for months while the inventor and attorney straighten it out—at considerable expense.

To make matters even more worrisome, the PTO proposal is based on software that is not yet available. The hoped-for program, called Author-In, is being developed at Intelligene (Mountain View, CA), the contractor that now operates Genbank. If Author-In works as envisioned, a scientist will be able to feed in a file of sequence data, which the program would then rearrange to fit the PTO's format—prompting the operator for each additional datum the PTO requires.

That, at least, is the hope, and PTO is betting heavily on it. The new rules are scheduled to assume the force of law on January 1, 1990—regardless of whether inventors and attorneys are ready to comply. It's hard to guess how much experience the officials at the PTO have actually had in working with computers, or whether they are aware of the reluctance of most scientists or computer specialists to invest in Version 1.0 of anything. It does seem painfully clear, however, that those officials are ignoring the PTO's own history and experience; according to the July 6 *Wall Street Journal*, the PTO's own overall computerization efforts have fallen three years behind schedule.

The PTO is undoubtedly pointing in the right direction; long sequences *should* be submitted in electronic form. It would be much more prudent, however, for the PTO to delay

implementing its rules until it has accumulated some practical experience with Author-In—or with whatever tool it chooses as the keystone of this policy.

Perhaps most important, the PTO should then phase in the new sequence requirements gradually, allowing a transition period during which it would accept ASCII sequence data in flexible formats, just as Genbank does. This would let examiners mount computer searches for prior art while the field adjusts to the new format and software.

The PTO should also train its own clerical people to convert computerized sequence data into the PTO format. It is easier and cheaper to thoroughly train a single clerk to use a piece of software daily than it is to shift that burden onto the shoulders of every biotech inventor and patent attorney—who likely will have to learn or re-learn Author-In every time they file a biotech patent (filings

"PTO officials are ignoring their own experience... the Office's own computerization efforts are three years behind schedule."

that can come months or even years apart for small companies, independent inventors, and academics). Sequence formatting by clerks who do it regularly would be more accurate, too, as the PTO knows well; patent examiners who work with applications filed under the Patent Cooperation Treaty (PCT) make frequent mistakes, because they see PCT-format applications only once or twice a year.

The PTO should be praised for its efforts. But at this crucial stage, a hasty misstep could turn into quite a messy fall.

—Patrick D. Kelly