Publication No. 661 (E)) on the procedures and requirements concerning the deposit of biological material, including procedures for obtaining a sample of deposited material, in each of the international depositary authorities.

(n) Indicate whether any special provisions, rules or procedures concerning electronic filing requirement for inventions dependent on nucleic or amino acid sequence information have been developed by the examining authority to address questions of enablement[14] or full description of any of the categories of inventions specified in question A.a). In their answers examining authorities should explain how such procedures operate and whether difficulties have been encountered in the practical application of those rules or procedures.

The PTO requires that applicants who have filed an application claiming an invention consisting of nucleotide or amino acid sequences provide a machine readable file containing the sequence information. Thus, where an applicant claims a particular protein or nucleotide sequence, and discloses that sequence in the application, the applicant will be required to provide the PTO with an electronic file containing the sequence information. Compliance with the rules govern submission of nucleotide or amino acid sequence information in electronic form (i.e., 37 CFR 1.821 et seq.) is a formalities, rather than substantive requirement. Failure to comply with the rules will result in an objection to the application form, rather than a substantive rejection based on inadequate disclosure.

Sections 2422 to 2432 of the MPEP explain how the PTO has implemented the rules governing sequence listing submissions. Useful overview sections have been reproduced below.

2420 The Requirements for Patent Applications Containing Nucleotide Sequence and/or Amino Acid Sequence Disclosures - the Sequence Rules

Prior to the effective date (October 1, 1990) and implementation of the sequence rules (37 CFR 1.821 through 1.825), applications for patents that included nucleotide or amino acid sequence information posed special problems for the Office. While not related to the disclosure requirements of an invention, problems existed in the presentation, examination and printing of nucleotide and amino acid sequence data that appeared in patent applications because of the lack of uniformity in submission of sequence data to the Office and the impracticality of properly searching and examining sequences submitted in paper form. In summary, the diversity and complexity of nucleotide and amino acid sequence data resulted in searching and analysis difficulties both within the Office and outside the Office, decreased accuracy of search and reproduction and increased costs. These difficulties made the development and implementation of the sequence rules a critical necessity for the Office. As such, the Office amended its regulations to establish a standardized format for descriptions of nucleotide and amino acid sequence data submitted as a part of patent applications, in conjunction with the required submission of that data in computer readable form. The final rules were published in the Federal Register at 55 FR 18230 (May 1, 1990) and in the Official Gazette at 1114 O.G. 29 (May 15, 1990). The sequence rules went into effect on October 1, 1990.

2421 Overview of the Sequence Rules

2421.01 Applications Affected

14.Enablement: A disclosure sufficiently clear and complete for the invention to be carried out by a person skilled in the art.

The sequence rules require the use of standard symbols and a standard format for sequence data in most sequence-type patent applications. They further require the submission of that data in computer readable form. Compliance is required for most disclosures of sequence data in new applications. For the purpose of the sequence rules, the term "new" with regard to applications means:

- For regular US applications, the application must have been filed on or after October 1, 1990. Continuing applications that claim a date prior to October 1, 1990, under 35 U.S.C. 120, except continuations-in-part (CIPs) filed on or after October 1, 1990, where material added includes a sequence, are not new applications.

- For PCT applications, the international filing date must be on or after October 1, 1990, and the US must be one of the designated states and the US must be the International Searching Authority and/or the International Preliminary Examining Authority. For national stage applications, the international filing date, not the 371 date, is used. For international applications, the international filing date, not the U.S. or foreign priority date, is used.

- Not applicable to reissues or reexams filed after October 1, 1990, unless application which matured into patent was also subject to rules.

- Not applicable to continuations or divisionals filed after October 1, 1990, unless parent application was also subject to rules.

- Will apply to CIPs where parent application subject to rules or where new material includes sequence information that falls within rules. Where a CIP is filed on or after October 1, 1990, and claims the benefit of a filing date prior to October 1, 1990, under 35 U.S.C. 120 and material added includes a sequence, the requirement for compliance is limited to the newly added sequence material. Full compliance, for all sequence information in the CIP is, however, encouraged.

The Office encourages voluntary compliance for applications not subject to rules, but all aspects of the rules must be complied with before data will be entered into the database. This includes submission of all statements required by the rules. In exceptional circumstances, it should be noted that the Office may waive the rules via a 37 CFR 1.183 petition.

2421.02 Summary of the Requirements of the Sequence Rules

Basically, the sequence rules define a set of symbols and procedures that are both mandatory and the only way that an applicant is permitted to describe information about a sequence that falls within the definitions used in the rules. Thus, 37 CFR 1.821 of the rules defines a sequence for the purpose of the rules, the requirements for specific symbols, formats, paper and computer readable copies of the sequence, and the deadlines for complying with the requirements. 37 CFR 1.822 to 37 CFR 1.824 set forth detailed descriptions of the requirements that are mandatory for the presentation of sequence data, and 37 CFR 1.825 sets forth procedures that are available to an applicant in the event that amendments to the sequence information or replacement of the computer readable copy become necessary.

The sequence rules embrace all unbranched nucleotide sequences with ten or more bases and all unbranched, non-D amino acid sequences with four or more amino acids. The rules apply to all sequences in a given application, whether claimed or not. All such sequences are relevant for the purposes of building a comprehensive database and properly assessing prior art. It is therefore essential that all sequences, whether only disclosed or also claimed, be included in the database.

2421.03 Notification of a Failure to Comply

With respect to the Office's determination of compliance with the sequence rules and the opportunities afforded applicants to satisfy the requirements of the rules, applicants will be notified of easily detectable deficiencies early in the application process. Deficiencies of a more sophisticated nature will likely only be detected by the examiner to whom the application is assigned. Applicants whose computer readable forms are damaged in the mail, are not readable, or are missing mandatory elements will be notified shortly after receipt of the application by the Office. Other errors or inconsistencies will be noted by the examiner early in the examination process. Upon detection of damage or a deficiency, a notice will be sent to the applicant detailing the damage or deficiency and setting at least a one month period for response. The period for response will usually be one month. However, if the notice is sent out with an Office communication having a longer period for response, the period for response may be longer than one month, e.g. where the notice is sent with an Office action on the merits setting a three month period for response. Extensions of time in which to reply will be available pursuant to 37 CFR 1.136. When an action by the applicant, such as a response to a notice to comply from the Office, is determined to be a bona fide attempt to comply with the rules and it is apparent that compliance with some requirement has inadvertently been omitted, the opportunity to explain and supply the omission will be given before the question of abandonment will be considered. See 37 CFR 1.135(c). The relevant form paragraphs and a copy of the Notice to Comply to be used in applications subject to the sequence rules are included in MPEP w 2427 Form Paragraphs and Notice to Comply.

A notification of a failure to comply with the sequence rules will usually be accompanied by an analysis of a submitted computer readable form. Any inquiries regarding a specific computer readable form that has been processed by the Office should be directed to the Systems Branch of the Biotechnology Division of the Scientific and Technical Information Center.

2429 Helpful Hints for Compliance

The Office has now had a good deal of experience in the implementation of the sequence rules. The following list sets forth helpful hints, for both examiners and applicants, for compliance. For the most part, the list is a compilation of frequently asked questions.

- Compliance is not a filing date issue.
- Compliance is not a 35 U.S.C. § 112 issue.
- Compliance is not a 35 U.S.C. 119/120 issue.
- Compliance is not a new matter issue. The standard for resolution of inconsistencies between the paper and the electronic copies and/or errors in the paper copy of sequence information is based on the new matter standard.

- Compliance can be achieved via amendment.

- The paper copy of Sequence Listing is an integral part of the application. The Sequence Listing should begin on a new page immediately prior to the claims, with pages numbered consecutively with the rest of the application. The new page that begins the "Sequence Listing" should be entitled "Sequence Listing." If not submitted as such at filing, the Sequence Listing must be inserted into the application via amendment, e.g., by preliminary amendment.

- Substitute pages must be used for changes to the Sequence Listing. - Punctuation, information item headings and parentheses listed in 37 CFR 1.823 are very important for our database. Extra colons and parentheses should not be used in Sequence Listings.

- The computer readable form cannot contain page numbers. Page numbers should only be placed on the paper copy of the Sequence Listing.

- The Patent In computer program is not the only means by which to comply with the rules. Any word processing program can be used to generate a Sequence Listing if it has the capability to convert a file into ASCII text. - If a word processing program is used to generate a "Sequence Listing," hard page break controls should not be used and margins should be adjusted to the smallest setting.

- However, word processing files should not be submitted to the Office; the Sequence Listing generated by a word processing file should be saved as an ASCII text file for submission. Most word processing programs provide this feature.

- Statements in accordance with 37 CFR 1.821(f), (g), (h) and 37 CFR 1.825 and proper labeling in accordance with 37 CFR 1.824(h) should be noted. Sample statements to support filings and submissions in accordance with 37 CFR 1.821 through 1.825 are provided in MPEP w 2428 Sample Statements.

- Use Box Sequence - 37 CFR 1.821(j).

- Three and a half inch disks are less fragile than five and a quarter inch disks.

- On nucleotide sequences, since only single strands may be depicted in the "Sequence Listing," please remember to show strands in 5' to 3' direction. - The single stranded nucleotide depicted in the "Sequence Listing" may represent a strand of a nucleotide sequence that may be single or double stranded which may be, further, linear or circular. An amino acid sequence or peptide may be linear or circular. In some instances, a sequence may be both single stranded and double stranded and/or both linear and circular. The response "not relevant" is also an acceptable response for both "Strandedness" and "Topology." - Current Application Data and Prior Application Data fields should always be supplied even if left blank. The Current Application Data headings should appear in the "Sequence Listing" in all cases. If the information about the current application is not known or is unavailable at the time of completing the Sequence Listing, then

the line following each heading should be left blank. This would normally be the case when the "Sequence Listing" is included in a newly filed application. Similarly, if information regarding prior applications is inapplicable or not known at the time of completing the "Sequence Listing" but will be later filed, then the Prior Application Data field headings should appear with the line following the heading left blank.

- Margin requirements and minimum character size for the Sequence Listing must be complied with and can usually be satisfied by using 10 point, 12 pitch type.

- If you receive a Notice to Comply that should not have been sent to you, a simple letter, in the form of a request for reconsideration of the notice, to the organization sending the notice should suffice to clarify the matter. - There are a limited number of Mandatory Items of Information. They are listed in MPEP w 2433 Sequence Listing Headings for Mandatory Items.

- Figures can be used to convey information not readily conveyed by the Sequence Listing. The exclusive conformance requirement of 37 CFR1.821(b) will be relaxed for drawing figures. However, the sequence information so conveyed must still be included in a "Sequence Listing" and the sequence identifier ("SEQ ID NO:X") must be used, either, in the drawing or in the "Brief Description of the Drawings."

- Extra copies of computer readable forms should not be sent to examiners. - Inosine should be represented by the use of "N".

- Stop codons, represented by an asterisk, are not permitted in amino acid sequences.

- Punctuation should not be used in a sequence to indicate unknown nucleotide bases or amino acid residues nor should punctuation be used to delimit active or functional regions of a sequence. These regions should be noted as Features of the sequence per 37 CFR 1.823(b)(2)(ix).

- The presence of an unnatural amino acid in a sequence does not have the same effect as the presence of a D-amino acid. The sequence may still be subject to the rules even though one or more of the amino acids is not naturally occurring. - Cyclic and branched peptides are causing some confusion in the application of the rules. Specific questions should be directed to Group 1800 personnel.

- A cyclic peptide with a tail is regarded as a branched sequence, and thereby exempt from the rules, if all bonds adjacent to the amino acid from which the tail emanates are normal peptide bonds.

- Representations of amino acid sequences, such as $XY(AA)_nZ$ where X, Y and Z can be various amino acids (properly formatted as Xaas) and (AA)n is a spacer where AA can be any amino acid and n can be 0, 1, 2