

ST.25	ST.26
ASCII .txt with numeric identifiers	XML with elements and attributes (WIPO's Sequence listing software available for download here: https://www.wipo.int/standards/en/sequence/index.html)
<p><u>Not required to include:</u></p> <p>D-amino acids</p> <p>Linear portions of branched sequences</p> <p>Nucleotide analogs</p>	<p><u>Must include:</u></p> <p>D-amino acids</p> <p>Linear portions of branched sequences</p> <p>Nucleotide analogs (e.g., PNAs and GNAs)</p>
<p><u>Annotation of sequences:</u></p> <p>Feature keys only</p>	<p><u>Annotation of sequences:</u></p> <p>Feature keys and qualifiers</p>
<p><u>Permitted to include sequences:</u></p> <p>< 10 specifically defined nucleotides</p> <p>< 4 specifically defined amino acids</p>	<p><u>Prohibited sequences:</u></p> <p>< 10 specifically defined nucleotides</p> <p>< 4 specifically defined amino acids</p> <p>**"Specifically defined" does not refer to "n" nucleotides or "X" amino acids</p>
ALL priority application information may be included	ONLY the earliest priority application can be included
ALL applicant and inventor names may be included	ONLY one applicant AND optionally ONE inventor may be included
One invention title permitted	Multiple invention titles permitted, each one in a different language
Applicant/inventor names and invention titles must be in basic Latin characters	Applicant/inventor names may be included using any valid Unicode character along with a basic Latin translation or transliteration
Sequences identified as DNA, RNA, or PRT only	Sequences identified as DNA, RNA, or AA along with a mandatory mol_type qualifier to further describe the molecule
<p><u>Organism names:</u></p> <p>Latin genus/species</p> <p>Virus name</p> <p>"artificial sequence"</p> <p>"unknown"</p>	<p><u>Organism names:</u></p> <ul style="list-style-type: none"> - Latin genus/species - Virus name - "synthetic construct" - "unidentified" <p>Common names, such as "mouse," must not be used as the organism name. If desired, common names can be included in the sequence listing in a note qualifier</p>
"u" represents uracil in nucleotide sequences	"t" represents uracil in RNA sequences and thymine in DNA sequences
Amino acid sequences represented by three letter abbreviations	Amino acid sequences represented by one letter abbreviations
"n" and "Xaa" variables must have a definition provided in a feature	Default value assumed for "n" and "X" variables with no definition
Feature location format not clearly defined	Strictly defined feature location formats; permits use of "" in all sequence types, and "A", "join", "order", and "complement" in nucleotide sequences
"Mixed mode" sequences permitted – nucleotide sequence with amino acid translation shown below	NO "mixed mode"; nucleotide translations are included in "translation" qualifiers only