



## NORME ST.25

### NORME RELATIVE À LA PRÉSENTATION DU LISTAGE DES SÉQUENCES DE NUCLÉOTIDES ET D'ACIDES AMINÉS DANS LES DEMANDES DE BREVET

*Révision adoptée par le Groupe de travail du SCIT sur les normes et la documentation à sa onzième session le 30 octobre 2009*

Il est recommandé que les offices appliquent les dispositions de la "Norme relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes internationales de brevet déposées selon le Traité de coopération en matière de brevets (PCT)", qui figure à l'annexe C des Instructions administratives du PCT, pour toutes les demandes de brevet autres que les demandes internationales déposées selon le PCT, sous réserve des modifications nécessaires compte tenu du fait que certaines dispositions propres aux procédures et prescriptions définies dans le cadre du PCT peuvent ne pas être applicables aux demandes de brevet autres que les demandes internationales déposées selon le PCT<sup>(\*)</sup>. Le texte de cette norme du PCT est reproduit ci-après.

---

<sup>(\*)</sup> Si, le 1<sup>er</sup> juillet 2009, la législation et la pratique nationales appliquées par un office ne sont pas compatibles avec les dispositions du paragraphe 3.i) de la "Norme relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes internationales de brevet déposées selon le Traité de coopération en matière de brevets (PCT)", qui stipulent qu'un listage des séquences figurant dans la demande telle qu'elle a été déposée 'doit être présenté dans une partie distincte de la description, être placé à la fin de la demande, s'intituler de préférence 'listage des séquences', commencer sur une nouvelle page et faire l'objet d'une pagination distincte", l'office peut choisir de ne pas appliquer ces dispositions tant que l'incompatibilité demeure.



## ANNEXE C

### NORME RELATIVE À LA PRÉSENTATION DES LISTAGES DES SÉQUENCES DE NUCLÉOTIDES ET D'ACIDES AMINÉS DANS LES DEMANDES INTERNATIONALES DE BREVET DÉPOSÉES SELON LE PCT

#### INTRODUCTION

1. La présente norme a été élaborée pour normaliser la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes internationales de brevet. Elle vise à permettre au déposant d'établir un listage unique qui soit acceptable pour tous les offices récepteurs, toutes les administrations chargées de la recherche internationale et les administrations chargées de l'examen préliminaire international aux fins de la phase internationale, ainsi que pour tous les offices désignés ou élus aux fins de la phase nationale. Elle vise aussi à accroître la précision et la qualité de la présentation des séquences de nucléotides et d'acides aminés dans les demandes internationales, à faciliter la présentation et la diffusion des séquences dans l'intérêt des déposants, du public et des examinateurs, à faciliter la recherche de données sur ces séquences ainsi qu'à permettre l'échange de données sur les séquences sous forme électronique et l'incorporation de ces données dans les bases de données informatisées.

#### DÉFINITIONS

2. Aux fins de la présente norme,

(i) l'expression "listage des séquences" désigne un listage des séquences de nucléotides ou d'acides aminés qui divulgue de façon détaillée les séquences de nucléotides ou d'acides aminés ainsi que d'autres informations disponibles;

*i-bis)* l'expression "listage des séquences faisant partie de la demande internationale" désigne un listage des séquences figurant dans la demande internationale telle qu'elle a été déposée (voir le paragraphe 3) ou un listage des séquences incorporé dans la demande internationale à la suite d'une correction en vertu de la règle 26, d'une rectification en vertu de la règle 91 ou d'une modification en vertu de l'article 34.2)b) de la description en rapport avec les séquences figurant dans la demande internationale telle qu'elle a été déposée (voir le paragraphe *3bis*);

*i-ter)* l'expression "listage des séquences ne faisant pas partie de la demande internationale" désigne un listage des séquences qui ne fait pas partie de la demande internationale mais a été remis aux fins de la recherche internationale ou de l'examen préliminaire international (voir les paragraphes 4 et *4bis*);

(ii) le terme "séquences" désigne des séquences linéaires d'au moins quatre acides aminés ou des séquences linéaires d'au moins dix nucléotides. Les séquences non linéaires, les séquences à moins de quatre nucléotides ou acides aminés spécialement définis, ainsi que les séquences comprenant des nucléotides ou des acides aminés différents de ceux qui sont énumérés dans les tableaux 1, 2, 3 et 4 de l'appendice 2 sont expressément exclues de cette définition;

(iii) le terme "nucléotides" désigne uniquement les nucléotides qui peuvent être représentés à l'aide des symboles indiqués dans le tableau 1 de l'appendice 2. Les modifications (par exemple, les bases méthylées) peuvent être décrites de la manière indiquée dans le tableau 2 de l'appendice 2, mais elles ne doivent pas être présentées de façon explicite dans la séquence de nucléotides;

(iv) le terme "acides aminés" désigne les acides aminés L que l'on rencontre généralement dans des protéines naturelles et qui sont énumérés dans le tableau 3 de l'appendice 2. N'entrent pas dans cette définition les séquences d'acides aminés qui contiennent au moins un acide aminé D. Toute



séquence d'acides aminés qui contient des acides aminés modifiés après traduction peut être représentée sous la forme de la séquence initialement traduite à l'aide des symboles indiqués dans le tableau 3 de l'appendice 2, les positions modifiées (par exemple hydroxylations ou glycosylations) étant elles-mêmes décrites de la manière indiquée dans le tableau 4 de l'appendice 2; ces modifications ne doivent toutefois pas être représentées de façon explicite dans la séquence d'acides aminés. Entre dans cette définition tout peptide ou toute protéine qui peut être exprimé(e) sous forme de séquence à l'aide des symboles énumérés dans le tableau 3 de l'appendice 2 et accompagné(e), par exemple, d'une description des liaisons anormales, des liaisons croisées (p. ex. ponts disulfure) et des coiffes terminales, des liaisons non peptidiques, etc.;

(v) l'expression "identificateur de séquence" désigne un nombre entier unique correspondant au SEQ ID NO attribué à chaque séquence figurant dans le listage;

(vi) l'expression "identificateur numérique" désigne un numéro à trois chiffres qui représente un élément de donnée déterminé;

(vii) l'expression "vocabulaire non connoté" désigne un vocabulaire contrôlé utilisé dans le listage des séquences qui représente des termes scientifiques de la façon prescrite par les fournisseurs de bases de données contenant des séquences (y compris des noms scientifiques, des qualificatifs et leur valeur en termes de vocabulaire contrôlé, les symboles figurant dans les tableaux 1, 2, 3 et 4 de l'appendice 2, et les clés de caractérisation, dans les tableaux 5 et 6 de l'appendice 2);

(viii) l'expression "administration compétente" désigne l'administration chargée d'effectuer la recherche internationale et d'établir l'opinion écrite de l'administration chargée de la recherche internationale pour la demande internationale en question, ou l'administration chargée d'effectuer l'examen préliminaire international pour la demande internationale en question.

## LISTAGE DES SÉQUENCES

### Listage des séquences faisant partie de la demande internationale

3. Un listage des séquences figurant dans la demande internationale telle qu'elle a été déposée doit :

i) être présenté dans une partie distincte de la description, être placé à la fin de la demande, de préférence être intitulé "Listage des séquences", commencer sur une nouvelle page et faire l'objet d'une pagination distincte<sup>34</sup>; de préférence, le listage des séquences ne doit pas être reproduit dans une autre partie de la demande; sous réserve du paragraphe 36, il n'est pas nécessaire de décrire les séquences ailleurs dans la description;

ii) présenter les séquences du listage des séquences ainsi que toute autre information devant figurer dans ce listage conformément aux paragraphes 5 à 35;

iii) lorsqu'il figure dans une demande internationale déposée sous forme électronique, être dans un format électronique de document et déposé selon un mode de transmission visé au paragraphe 37.

*3bis.* Toute correction en vertu de la règle 26, toute rectification en vertu de la règle 91 ou toute modification en vertu de l'article 34.2)b) de la description qui est remise en rapport avec des séquences figurant dans une demande internationale déposée sous forme électronique doit être remise sous la forme d'un listage des séquences sous forme électronique contenant l'intégralité du listage et présentant la correction, la rectification ou la modification pertinente. Ce listage des séquences doit :

<sup>34</sup> *Note de l'éditeur* : Il n'est pas nécessaire que le listage des séquences fasse l'objet d'une pagination distincte lorsqu'il figure dans une demande internationale déposée sous forme électronique et qu'il est dans le format électronique de document visé au paragraphe 40.



i) de préférence être intitulé “Listage des séquences – Correction”, “Listage des séquences – Rectification” ou “Listage des séquences – Modification”, selon le cas, et faire l’objet d’une pagination distincte<sup>34</sup>;

ii) présenter les séquences du listage des séquences ainsi que toute autre information devant figurer dans ce listage conformément aux paragraphes 5 à 35; le cas échéant, la numérotation originale des séquences figurant dans la demande internationale telle qu’elle a été déposée (voir le paragraphe 5) doit être maintenue; à défaut, les séquences doivent être numérotées conformément au paragraphe 5;

iii) être dans un format électronique de document et déposé selon un mode de transmission visé au paragraphe 38.

### Listage des séquences ne faisant pas partie de la demande internationale

4. Un listage des séquences fourni en vertu de la règle 13<sup>ter</sup> aux fins de la recherche internationale ou de l’examen préliminaire international doit :

i) de préférence être intitulé “Listage des séquences – Règle 13<sup>ter</sup>”;

ii) présenter les séquences du listage des séquences ainsi que tous autres renseignements devant figurer dans ce listage conformément aux paragraphes 5 à 35; le cas échéant, la numérotation originale des séquences figurant dans la demande internationale telle qu’elle a été déposée (voir le paragraphe 5) doit être maintenue; à défaut, les séquences doivent être numérotées conformément au paragraphe 5;

iii) lorsqu’il est fourni sur papier conformément à la règle 13<sup>ter</sup>.1.b), faire l’objet d’une pagination distincte;

iv) lorsqu’il est fourni sous forme électronique, être dans un format électronique de document et déposé selon un mode de transmission visé au paragraphe 39;

v) lorsqu’il est fourni sous forme électronique en même temps que la demande internationale, être identique au listage des séquences figurant dans la demande et être accompagné d’une déclaration selon laquelle “les informations enregistrées sous forme électronique fournies en vertu de la règle 13<sup>ter</sup> sont identiques à celles du listage des séquences figurant dans la demande internationale”;

vi) lorsqu’il est fourni après le dépôt de la demande internationale, ne pas contenir d’éléments allant au delà de la divulgation faite dans la demande internationale telle qu’elle a été déposée et être accompagné d’une déclaration dans ce sens; ce listage des séquences ne doit contenir que les séquences qui ont été divulguées dans la demande internationale telle qu’elle a été déposée.

4bis. Toute correction en vertu de la règle 26, toute rectification en vertu de la règle 91 ou toute modification en vertu de l’article 34.2)b) de la description qui est remise en rapport avec des séquences figurant dans la demande internationale telle qu’elle a été déposée doit être accompagnée, aux fins de la recherche internationale ou de l’examen préliminaire international, d’un listage des séquences sous forme électronique, dans un format électronique de document conformément au paragraphe 39, contenant l’intégralité du listage, y compris la correction, la rectification ou la modification en question, chaque fois que l’administration compétente l’exige, à moins que cette administration n’ait déjà accès à ce listage sous forme électronique sous une forme et d’une manière qu’elle accepte. Ledit listage des séquences sous forme électronique doit :

i) de préférence être intitulé “Listage des séquences – Correction – Règle 13<sup>ter</sup>”, “Listage des séquences – Rectification – Règle 13<sup>ter</sup>” ou “Listage des séquences – Modification – Règle 13<sup>ter</sup>”, selon le cas;

ii) présenter les séquences du listage des séquences ainsi que toute autre information devant figurer dans ce listage conformément aux paragraphes 5 à 35; le cas échéant, la



numérotation originale des séquences figurant dans la demande internationale telle qu'elle a été déposée (voir le paragraphe 5) doit être maintenue; à défaut, les séquences doivent être numérotées conformément au paragraphe 5;

- iii) être déposé selon un mode de transmission visé au paragraphe 39;
- iv) être identique au listage des séquences remis, conformément au paragraphe 3*bis*, sous forme de correction en vertu de la règle 26, de rectification en vertu de la règle 91 ou de modification en vertu de l'article 34.2.b) de la description et être accompagné d'une déclaration selon laquelle "les informations enregistrées sous forme électronique fournies en vertu de la règle 13*ter* sont identiques à celles du listage des séquences remis sous forme de correction en vertu de la règle 26 (ou de rectification en vertu de la règle 91, ou de modification en vertu de l'article 34.2.b), selon le cas) de la description en rapport avec des séquences figurant dans la demande internationale telle qu'elle a été déposée".

Lorsque l'administration compétente n'a pas accès au listage des séquences sous forme électronique et, le cas échéant, à la déclaration y relative, elle ne doit prendre en considération la correction, rectification ou modification en question aux fins de la recherche internationale ou de l'examen préliminaire international que dans la mesure où elle peut effectuer une recherche ou un examen préliminaire significatifs sans disposer de ce listage des séquences sous forme électronique.

## PRÉSENTATION DES SÉQUENCES

5. À chaque séquence doit être attribué un identificateur de séquence distinct, la progression numérique étant séquentielle et commençant à 1. Lorsque aucune séquence ne figure sous l'identificateur de séquence, le code 000 doit apparaître sous l'identificateur numérique <400>, en commençant sur la ligne qui suit la mention SEQ ID NO. En regard de l'identificateur numérique <160> doit être indiqué le nombre total de séquences, suivi d'une séquence ou du code 000.

6. Dans la description, les revendications ou les dessins de la demande, toute séquence du listage à laquelle il est fait référence doit être désignée par son identificateur de séquence et précédée de la mention "SEQ ID NO:".

7. Les séquences de nucléotides et d'acides aminés doivent être représentées au moins sous l'une des trois formes suivantes :

- i) une séquence de nucléotides pure;
- ii) une séquence d'acides aminés pure;
- iii) une séquence de nucléotides et sa séquence d'acides aminés correspondante.

En ce qui concerne les séquences divulguées selon le mode de présentation indiqué sous iii), la séquence d'acides aminés doit être divulguée séparément dans le listage des séquences en tant que séquence d'acides aminés pure assortie d'un identificateur de séquence distinct.

## Séquences de nucléotides

### *Symboles à utiliser*

8. Toute séquence de nucléotides doit être représentée par un seul brin de codage, dans le sens 5'-3' et de gauche à droite. Les valeurs 3' et 5' ne doivent pas être représentées dans la séquence.

9. Les bases d'une séquence de nucléotides doivent être représentées au moyen du code à une lettre utilisé pour les séquences de ce type. Seules les lettres minuscules indiquées dans le tableau 1 de l'appendice 2 doivent être utilisées.



10. Les bases modifiées doivent être représentées par les bases non modifiées correspondantes contenues dans la séquence elle-même (ou au moyen du symbole “n”) lorsque la base modifiée figure parmi celles qui sont énumérées dans le tableau 2 de l’appendice 2; la modification doit faire l’objet d’une description plus détaillée dans la section “caractéristiques” du listage des séquences, à l’aide des codes énumérés dans le tableau 2 de l’appendice 2. Ces codes peuvent être utilisés dans la description ou dans la partie “caractéristiques” du listage des séquences mais pas dans la séquence proprement dite (voir aussi le paragraphe 32). Le symbole “n” est l’équivalent d’un seul nucléotide inconnu ou modifié.

#### *Mode de présentation à suivre*

11. Une séquence de nucléotides doit comporter 60 bases par ligne au maximum, avec un espace entre chaque codon ou groupe de 10 bases.

12. Les bases d’une séquence de nucléotides (y compris les introns) doivent figurer sur la liste par groupes de 10 bases, sauf celles qui se situent dans les régions codantes de la séquence. Les bases (moins de 10) qui restent à l’extrémité des régions non codantes d’une séquence doivent être regroupées et séparées des groupes voisins par un espace.

13. Les bases des régions codantes d’une séquence de nucléotides doivent figurer sur la liste sous forme de triplets (codons).

14. L’énumération des nucléotides doit commencer par la première base de la séquence, qui portera le numéro 1. Elle doit être continue dans toute la séquence dans le sens 5’-3’. Elle doit figurer dans la marge de droite sur la ligne contenant les codes à une lettre correspondant aux bases et indiquer le numéro de la dernière base de cette ligne. La méthode présentée ci-dessus pour énumérer des séquences de nucléotides s’applique aussi aux séquences de nucléotides de configuration circulaire, à cette différence près que la désignation du premier nucléotide de la séquence peut être laissée au choix du déposant.

15. Toute séquence composée d’un segment ou de plusieurs segments non contigus d’une séquence plus grande ou de segments provenant de différentes séquences doit être numérotée comme une séquence distincte, au moyen d’un identificateur de séquence distinct. Une séquence comportant un ou des espaces doit être numérotée comme une série de séquences distinctes, au moyen d’identificateurs distincts, le nombre de séquences distinctes étant égal au nombre de chaînes continues.

#### **Séquences d’acides aminés**

##### *Symboles à utiliser*

16. Les acides aminés d’une séquence protéique ou peptidique doivent être énumérés dans le sens amino-carboxy et de gauche à droite. Les groupes amino et carboxy ne doivent pas être représentés dans la séquence.

17. Les acides aminés doivent être représentés au moyen du code à trois lettres (la première lettre étant en majuscule), conformément à la liste qui figure dans le tableau 3 de l’appendice 2. Une séquence d’acides aminés qui contient un espace ou des symboles internes de fin (par exemple, “Ter,” “\*” ou “.”) ne peut pas être représentée comme une séquence d’acides aminés unique, mais doit être présentée comme une séquence d’acides aminés distincte (voir le paragraphe 22).

18. Les acides aminés modifiés et peu connus doivent être représentés comme les acides aminés non modifiés correspondants dans la séquence elle-même (ou par “Xaa”) si l’acide aminé modifié figure parmi ceux qui sont énumérés dans le tableau 4 de l’appendice 2; la modification doit faire l’objet



d'une description plus détaillée dans la section "caractéristiques" du listage des séquences, à l'aide des codes énumérés dans le tableau 4 de l'appendice 2. Ces codes peuvent être utilisés dans la description ou dans la partie "caractéristiques" du listage des séquences mais pas dans la séquence proprement dite (voir aussi le paragraphe 32). Le symbole "Xaa" est l'équivalent d'un seul acide aminé inconnu ou modifié.

*Mode de présentation à suivre*

19. Une séquence de protéines ou de peptides doit comporter 16 acides aminés par ligne au maximum, avec un espace entre chaque acide aminé.

20. Les acides aminés correspondant aux codons dans les régions codantes d'une séquence de nucléotides doivent figurer immédiatement sous les codons correspondants. Lorsqu'un codon est scindé par un intron, le symbole d'acide aminé doit figurer sous la partie du codon contenant deux nucléotides.

21. L'énumération des acides aminés doit commencer par le premier acide aminé de la séquence, qui portera le numéro 1. À titre facultatif, les acides aminés précédant la protéine mature, par exemple les préséquences, les proséquences et les pré-proséquences ainsi que les séquences signal lorsqu'elles existent, doivent porter des nombres négatifs et être numérotés à rebours, en commençant par l'acide aminé voisin de l'acide portant le numéro 1. Le numéro 0 (zéro) n'est pas utilisé lorsque, dans la numérotation des acides aminés, des valeurs négatives servent à distinguer la protéine mature. Le numéro doit figurer dans la séquence tous les cinq acides. La méthode présentée ci-dessus pour énumérer les séquences d'acides aminés s'applique aussi aux séquences d'acides aminés de configuration circulaire, si ce n'est que la désignation du premier acide aminé de la séquence peut être laissée au choix du déposant.

22. Toute séquence d'acides aminés composée d'un segment ou de plusieurs segments non contigus d'une séquence plus grande ou de segments provenant de différentes séquences doit être numérotée comme une séquence distincte, au moyen d'un identificateur de séquence distinct. Une séquence comportant un ou des espaces doit être numérotée comme une série de séquences distinctes, au moyen d'identificateurs de séquence distincts, le nombre de séquences distinctes étant égal au nombre de chaînes continues.

**AUTRES INFORMATIONS DEVANT FIGURER DANS LE LISTAGE DES SÉQUENCES**

23. L'ordre des éléments d'information dans le listage des séquences doit suivre l'ordre dans lequel ces éléments sont énumérés dans la liste des identificateurs numériques des éléments de données définis à l'appendice 1.

24. Seuls les identificateurs numériques des éléments de données définis à l'appendice 1 peuvent être utilisés aux fins de la présentation des éléments d'information figurant dans le listage des séquences. Les descriptions correspondantes des identificateurs numériques ne doivent pas être utilisées. Les renseignements communiqués doivent suivre immédiatement l'identificateur numérique et seuls les identificateurs numériques pour lesquels des renseignements sont communiqués doivent figurer dans le listage des séquences. Font exception à cette règle les identificateurs numériques <220> et <300>, qui servent d'en-tête aux éléments "caractéristique" et "informations concernant la publication" et sont associés à l'information figurant sous les identificateurs numériques <221> à <223> et <301> à <313>, respectivement. Lorsque, sous ces identificateurs numériques, des indications concernant les éléments "caractéristique" et "informations concernant la publication" sont fournies dans le listage des séquences, il convient d'inclure les identificateurs numériques <220> et <300> en les laissant en blanc. En général, une ligne vierge doit être insérée entre les identificateurs numériques lorsque le chiffre venant en première ou en deuxième position dans ces identificateurs change. Toutefois, aucune



ligne vierge ne doit précéder l'identificateur numérique <310>. De plus, une ligne vierge doit précéder tout identificateur numérique répété.

*Éléments de données obligatoires*

25. Le listage des séquences doit contenir, en sus de la séquence de nucléotides ou d'acides aminés proprement dite et juste avant celle-ci, les éléments d'information ci-après définis à l'appendice 1 (éléments de données obligatoires) :

<110>	Nom du déposant
<120>	Titre de l'invention
<160>	Nombre de SEQ ID NO
<210>	SEQ ID NO : x
<211>	Longueur
<212>	Type
<213>	Organisme
<400>	Séquence

Lorsque le nom du déposant (identificateur numérique <110>) est écrit dans des caractères qui ne sont pas ceux de l'alphabet latin, il convient de l'écrire aussi à l'aide de cet alphabet, sous la forme d'une simple translittération ou d'une traduction en anglais.

Tous les éléments de données, à l'exception de ceux qui figurent sous les identificateurs numériques <110>, <120> et <160>, doivent être répétés pour chaque séquence figurant dans le listage des séquences. Seuls sont obligatoires les éléments de données figurant sous les identificateurs numériques <210> et <400> lorsqu'aucune séquence n'est associée à un identificateur de séquence (voir le paragraphe 5 ci-dessus et la séquence SEQ ID NO 4 dans l'exemple donné à l'appendice 3 de la présente norme).

26. Outre les éléments de données indiqués au paragraphe 25, lorsqu'un listage des séquences est fourni à un moment quelconque avant l'attribution d'un numéro de demande, l'élément de donnée ci-après doit être inclus dans le listage des séquences :

<130>	Référence du dossier
-------	----------------------

27. Outre les éléments de données indiqués au paragraphe 25, lorsqu'un listage des séquences est fourni à un moment quelconque après l'attribution d'un numéro de demande, les éléments de données ci-après doivent être inclus dans le listage des séquences :

<140>	Demande de brevet actuelle
<141>	Date de dépôt de la demande actuelle

28. Outre les éléments de données indiqués au paragraphe 25, lorsqu'un listage des séquences est déposé en relation avec une demande dans laquelle une priorité est revendiquée, les éléments de données ci-après doivent être inclus dans le listage des séquences :

<150>	Demande de brevet antérieure
<151>	Date de dépôt de la demande antérieure





29. Lorsque “n”, “Xaa”, une base modifiée ou un acide aminé L modifié ou peu connu est utilisé dans la séquence, les éléments de données ci-après doivent être inclus :

<220>	Caractéristique
<221>	Nom/clé
<222>	Emplacement
<223>	Autres informations

30. Lorsque l'organisme (identificateur numérique <213>) est une “séquence artificielle” ou qu'il est “inconnu”, les éléments de données ci-après doivent être inclus :

<220>	Caractéristique
<223>	Autres informations

#### *Éléments de données facultatifs*

31. Tous les éléments de données définis à l'appendice 1 qui ne sont pas mentionnés dans les paragraphes 25 à 30 sont facultatifs (éléments de données facultatifs).

#### *Présentation des caractéristiques*

32. Les caractéristiques des séquences (identificateur numérique <220>) doivent être décrites à l'aide des “clés de caractérisation” indiquées dans les tableaux 5 et 6 de l'appendice 2<sup>35</sup>.

#### *Texte libre*

33. Par “texte libre” on entend la description des caractéristiques d'une séquence dans le cadre de l'identificateur numérique 223 (autres informations) à l'aide d'un vocabulaire qui ne fait pas partie du vocabulaire non connoté défini au paragraphe 2)vii).

34. Le texte libre doit se limiter à quelques termes brefs indispensables à la compréhension de la séquence. Il ne doit pas excéder quatre lignes, avec un maximum de 65 caractères par ligne, pour chaque élément de données lorsqu'il est écrit en anglais. Toute autre information doit figurer dans la partie principale de la description dans la langue de celle-ci.

35. Le texte libre doit, de préférence, être rédigé en anglais.

### **RÉPÉTITION DU TEXTE LIBRE DANS LA PARTIE PRINCIPALE DE LA DESCRIPTION**

36. Lorsque le listage des séquences faisant partie de la demande internationale contient du texte libre, celui-ci doit être répété dans la partie principale de la description, dans la même langue. Il est recommandé que le texte libre figurant dans la langue de la partie principale de la description soit inséré dans une rubrique particulière de la description intitulée “Texte libre du listage des séquences”.

### **LISTAGES DES SÉQUENCES SOUS FORME ÉLECTRONIQUE**

37. Tout listage des séquences visé au paragraphe 3 figurant dans une demande internationale déposée sous forme électronique doit être dans un format électronique de document et déposé selon un mode de transmission indiqué par l'office récepteur aux fins du dépôt des demandes internationales sous

<sup>35</sup> *Note de l'éditeur* : Ces tableaux contiennent des extraits du DDBJ/EMBL/GenBank Feature Table (séquences de nucléotides) et du tableau de caractéristiques SWISS PROT (séquences d'acides aminés).



forme électronique, étant entendu que ce listage doit être de préférence dans le format électronique de document indiqué au paragraphe 40 et, si possible, déposé selon un mode de transmission indiqué aussi bien par l'office récepteur que par l'administration compétente.<sup>36, 37</sup>

38. Tout listage des séquences sous forme électronique visé au paragraphe 3bis doit être dans un format électronique de document qui a été indiqué par l'office récepteur (s'il s'agit d'une correction) ou par l'administration compétente (s'il s'agit d'une rectification ou d'une modification) aux fins du dépôt des demandes internationales sous forme électronique, étant entendu que ce listage doit être de préférence dans le format électronique de document indiqué au paragraphe 40 et, si possible, déposé selon un mode de transmission indiqué aussi bien par l'office récepteur que par l'administration compétente. Ce listage doit être déposé selon un mode de transmission indiqué par l'office récepteur ou l'administration compétente, selon le cas, aux fins du présent paragraphe; si possible, il doit être déposé de préférence selon un mode de transmission indiqué aussi bien par l'office récepteur que par l'administration compétente.<sup>38</sup>

39. Tout listage des séquences sous forme électronique visé aux paragraphes 4 et 4bis fourni aux fins de la recherche internationale ou de l'examen préliminaire international doit être dans le format électronique de document indiqué au paragraphe 40 et déposé selon un mode de transmission qui a été indiqué par l'administration compétente aux fins du présent paragraphe.

40. Aux fins de la recherche internationale et de l'examen préliminaire international, tout listage des séquences sous forme électronique doit figurer dans un seul fichier électronique codé comme un fichier texte selon la page de code IBM<sup>39</sup> 437, la page de code IBM 932<sup>40</sup> ou une page de code compatible de manière à représenter le listage des séquences conformément aux dispositions des paragraphes 5 à 36 sans aucun autre code. Une page de code compatible (requis pour les caractères japonais, chinois, cyrilliques, arabes, grecs, hébraïques, etc.) est une page de code qui attribue les lettres de l'alphabet romain et les chiffres aux mêmes positions hexadécimales que les pages de code indiquées.

41. Tout listage des séquences dans le format électronique de document indiqué au paragraphe 40 doit, de préférence, être créé par un logiciel spécialisé tel que PatentIn.

<sup>36</sup> *Note de l'éditeur* : Lorsqu'un listage des séquences sous forme électronique conforme à la présente norme n'est pas fourni à l'administration compétente sous une forme et d'une manière qu'elle accepte (à savoir, en particulier, lorsqu'il ne lui est pas fourni dans le format électronique de document visé au paragraphe 40), l'administration compétente peut inviter le déposant à lui fournir ce listage des séquences sous forme électronique (voir la règle 13ter).

<sup>37</sup> *Note de l'éditeur* : Quel que soit le format électronique de document dans lequel le listage des séquences est présenté, l'agencement (par exemple, colonnes et rangées) des éléments de données inclus dans le listage des séquences et le format des séquences de nucléotides ou d'acides aminés proprement dites indiqués dans la présente annexe doivent être conservés.

<sup>38</sup> *Note de l'éditeur* : Lorsqu'un listage des séquences de remplacement sous forme électronique, contenant une correction, une rectification ou une modification, n'est pas fourni à l'administration compétente sous une forme et d'une manière qu'elle accepte (à savoir, en particulier, lorsqu'il ne lui est pas fourni dans le format électronique de document indiqué au paragraphe 40), ladite administration ne doit prendre en considération cette correction, rectification ou modification aux fins de la recherche internationale ou de l'examen préliminaire international que dans la mesure où elle peut effectuer une recherche significative ou un examen préliminaire significatif sans disposer de ce listage des séquences de remplacement (voir le paragraphe 4bis). Voir aussi la note de l'éditeur 39, qui s'applique également à tout listage des séquences de remplacement sous forme électronique visé au paragraphe 3bis.

<sup>39</sup> *Note de l'éditeur* : IBM est une marque enregistrée de la International Business Machines Corporation des États-Unis d'Amérique.

<sup>40</sup> *Note de l'éditeur* : Les pages de code mentionnées constituent des normes *de facto* pour les ordinateurs personnels.



## PROCÉDURE DEVANT LES OFFICES DÉSIGNÉS ET ÉLUS

42. Aux fins de la procédure devant un office désigné ou élu au sein duquel le traitement d'une demande internationale contenant la divulgation d'une ou de plusieurs séquences de nucléotides ou d'acides aminés a commencé (voir la règle 13<sup>ter</sup>.3),

i) toute mention de l'office récepteur ou de l'administration compétente s'entend comme une mention de l'office désigné ou élu concerné;

ii) toute mention d'un listage des séquences incorporé dans la demande internationale sous forme de correction en vertu de la règle 26, de rectification en vertu de la règle 91 ou de modification en vertu de l'article 34.2)b) de la description en rapport avec les séquences figurant dans la demande telle qu'elle a été déposée est à interpréter comme visant aussi tout listage des séquences inclus dans la demande, en vertu de la législation nationale appliquée par l'office désigné ou élu concerné, sous forme de correction (d'une irrégularité de forme), de rectification (d'une erreur évidente) ou de modification de la description en rapport avec les séquences figurant dans la demande telle qu'elle a été déposée;

iii) toute mention d'un listage des séquences fourni aux fins de la recherche internationale ou de l'examen préliminaire international est à interpréter comme visant aussi tout listage des séquences fourni à l'office désigné ou élu concerné aux fins de la recherche nationale ou de l'examen national par cet office;

iv) l'office désigné ou élu concerné peut inviter le déposant à lui fournir, dans un délai raisonnable en l'espèce, aux fins de la recherche nationale ou de l'examen national, un listage des séquences sous forme électronique conforme à la présente norme, à moins que l'office n'ait déjà accès à ce listage sous forme électronique sous une forme et d'une manière qu'il accepte.

### Appendices

*Appendice 1 : Identificateurs numériques*

*Appendice 2 : Symboles des nucléotides et des acides aminés et tableau des caractéristiques*

Tableau 1 : liste des nucléotides

Tableau 2 : liste des nucléotides modifiés

Tableau 3 : liste des acides aminés

Tableau 4 : liste des acides aminés modifiés ou peu connus

Tableau 5 : liste des clés de caractérisation concernant les séquences de nucléotides

Tableau 6 : liste des clés de caractérisation concernant les séquences de protéines

*Appendice 3 : Exemple de listage des séquences*

[Les appendices 1 à 3 de l'annexe C suivent]



### Annexe C, appendice 1 Identificateurs numériques

Seuls les identificateurs numériques définis ci-après peuvent être utilisés dans les listages de séquences soumis avec les demandes. Le texte des en-têtes d'éléments de données indiqués ne doit pas figurer dans les listages.

Les identificateurs numériques des éléments de données obligatoires, c'est-à-dire les éléments de données qui doivent impérativement figurer dans les listages de séquences (voir le paragraphe 25 de la présente norme : identificateurs 110, 120, 160, 210, 211, 212, 213 et 400), et les identificateurs numériques des éléments de données qui doivent impérativement être indiqués dans les cas mentionnés dans la présente norme (voir les paragraphes 26, 27, 28, 29 et 30 de la présente norme : identificateurs 130, 140, 141, 150 et 151 et 220 à 223), sont signalés par la lettre "O".

Les identificateurs numériques des éléments de données facultatifs (voir le paragraphe 31 de la présente norme) sont signalés par la lettre "F".

Identificateur numérique	Description de l'identificateur numérique	Obligatoire (O) ou facultatif (F)	Observations
<110>	Nom du déposant	O	lorsque le nom du déposant est écrit dans des caractères qui ne sont pas ceux de l'alphabet latin, il convient de l'écrire à l'aide de cet alphabet, sous la forme d'une simple translittération ou d'une traduction en anglais
<120>	Titre de l'invention	O	
<130>	Référence du dossier	O, dans les cas mentionnés au paragraphe 26 de la présente norme	voir le paragraphe 26 de la présente norme
<140>	Demande de brevet actuelle	O, dans les cas mentionnés au paragraphe 27 de la présente norme	voir le paragraphe 27 de la présente norme; la demande de brevet actuelle doit être désignée au moyen des éléments ci-après, dans l'ordre qui suit : le code à deux lettres indiqué conformément à la norme ST.3 de l'OMPI et le numéro de la demande (sous le format de numérotation utilisé par l'office de propriété industrielle auprès duquel la demande de brevet actuelle est déposée) ou, pour une demande internationale, le numéro de demande internationale



## MANUEL SUR L'INFORMATION ET LA DOCUMENTATION EN MATIÈRE DE PROPRIÉTÉ INDUSTRIELLE

Réf. : Normes – ST.25

page : 3.25.13

[Annexe C, appendice 1, suite]

Identificateur numérique	Description de l'identificateur numérique	Obligatoire (O) ou facultatif (F)	Observations
<141>	Date de dépôt de la demande actuelle	O, dans les cas mentionnés au paragraphe 27 de la présente norme	voir le paragraphe 27 de la présente norme; la date doit être indiquée conformément à la norme ST.2 de l'OMPI (CCYY MM DD)
<150>	Demande de brevet antérieure	O, dans les cas mentionnés au paragraphe 28 de la présente norme	voir le paragraphe 28 de la présente norme; la demande de brevet antérieure doit être désignée au moyen des éléments ci-après, dans l'ordre qui suit : le code à deux lettres indiqué conformément à la norme ST.3 de l'OMPI, et le numéro de la demande (sous le format de numérotation utilisé par l'office de propriété industrielle auprès duquel la demande de brevet antérieure a été déposée) ou, pour une demande internationale, le numéro de demande internationale
<151>	Date de dépôt de la demande antérieure	O, dans les cas mentionnés au paragraphe 28 de la présente norme	voir le paragraphe 28 de la présente norme; la date doit être indiquée conformément à la norme ST.2 de l'OMPI (CCYY MM DD)
<160>	Nombre de SEQ ID NO	O	
<170>	Logiciel	F	
<210>	Information concernant SEQ ID NO : x	O	nombre entier représentant la séquence SEQ ID NO
<211>	Longueur	O	la longueur de la séquence est exprimée en nombre de paires de bases ou d'acides aminés
<212>	Type	O	type de molécule séquencée dans SEQ ID NO: x, soit ADN, soit ARN, soit PRT; si une séquence de nucléotides contient à la fois des fragments d'ADN et des fragments d'ARN, la valeur sera "ADN"; en outre, la molécule combinée ADN/ARN sera décrite également dans les identificateurs <220> à <223>, sous Caractéristique
<213>	Organisme	O	espèce (nom scientifique) ou "Séquence artificielle" ou "Non connu"



## MANUEL SUR L'INFORMATION ET LA DOCUMENTATION EN MATIÈRE DE PROPRIÉTÉ INDUSTRIELLE

Réf. : Normes – ST.25

page : 3.25.14

[Annexe C, appendice 1, suite]

Identificateur numérique	Description de l'identificateur numérique	Obligatoire (O) ou facultatif (F)	Observations
<220>	Caractéristique	O, dans les cas mentionnés aux paragraphes 29 et 30 de la présente norme	laisser en blanc; voir les paragraphes 29 et 30 de la présente norme; description des points ayant une importance biologique dans la séquence SEQ ID NO: x (peut être répétée en fonction du nombre des caractéristiques indiquées)
<221>	Nom/clé	O, dans les cas mentionnés au paragraphe 29 de la présente norme	voir le paragraphe 29 de la présente norme; seules doivent être utilisées les clés décrites dans les tableaux 5 ou 6 de l'appendice 2
<222>	Emplacement	O, dans les cas mentionnés au paragraphe 29 de la présente norme	voir le paragraphe 29 de la présente norme; – de (numéro de la première base ou du premier acide aminé dans la caractéristique) – à (numéro de la dernière base ou du dernier acide aminé dans la caractéristique) – paires de bases (numéros des positions des paires de bases dans une séquence de nucléotides) – acides aminés (numéros des positions des résidus d'acides aminés dans une séquence d'acides aminés) – lorsque la caractéristique se situe sur un brin complémentaire de celui figurant dans le listage des séquences



## MANUEL SUR L'INFORMATION ET LA DOCUMENTATION EN MATIÈRE DE PROPRIÉTÉ INDUSTRIELLE

Réf. : Normes – ST.25

page : 3.25.15

[Annexe C, appendice 1, suite]

Identificateur numérique	Description de l'identificateur numérique	Obligatoire (O) ou facultatif (F)	Observations
<223>	Autres informations :	O, dans les cas mentionnés aux paragraphes 29 et 30 de la présente norme	voir les paragraphes 29 et 30 de la présente norme; toute autre information pertinente, exprimée à l'aide du vocabulaire non connoté ou sous forme de texte libre (de préférence en anglais); tout texte libre doit être répété dans la partie principale de la description, dans la même langue (voir le paragraphe 36 de la présente norme); lorsqu'une base modifiée ou un acide aminé modifié ou peu connu figurant dans les tableaux 2 et 4 de l'appendice 2 est utilisé dans la séquence, il convient d'utiliser le symbole associé à cette base ou à cet acide aminé des tableaux 2 et 4 de l'appendice 2
<300>	Informations concernant la publication	F	laisser en blanc; à répéter pour chaque publication pertinente
<301>	Auteurs	F	
<302>	Titre	F	titre de la publication
<303>	Périodique	F	périodique dans lequel les données ont été publiées
<304>	Volume	F	volume du bulletin officiel dans lequel les données ont été publiées
<305>	Numéro	F	numéro du périodique dans lequel les données ont été publiées
<306>	Pages	F	numéro des pages du périodique dans lequel les données ont été publiées
<307>	Date	F	date de parution du périodique dans lequel les données ont été publiées; si possible, la date doit être indiquée conformément à la norme ST.2 de l'OMPI (CCYY MM DD)
<308>	Numéro d'entrée dans la base de données	F	numéro d'entrée attribué par la base de données, y compris nom de cette base de données
<309>	Date d'entrée dans la base de données	F	date d'entrée dans la base de données; la date doit être indiquée conformément à la norme ST.2 de l'OMPI (CCYY MM DD)



## MANUEL SUR L'INFORMATION ET LA DOCUMENTATION EN MATIÈRE DE PROPRIÉTÉ INDUSTRIELLE

Réf. : Normes – ST.25

page : 3.25.16

[Annexe C, appendice 1, suite]

Identificateur numérique	Description de l'identificateur numérique	Obligatoire (O) ou facultatif (F)	Observations
<310>	Numéro du document	F	numéro du document, uniquement pour les citations de brevets; le document complet doit comprendre, dans l'ordre : le code à deux lettres indiqué conformément à la norme ST.3 de l'OMPI, le numéro de publication indiqué conformément à la norme ST.6 de l'OMPI et le code de type de document indiqué conformément à la norme ST.16 de l'OMPI
<311>	Date de dépôt	F	date de dépôt du document, uniquement pour les citations de brevets; la date doit être indiquée conformément à la norme ST.2 de l'OMPI (CCYY MM DD)
<312>	Date de publication	F	date de publication du document, uniquement pour les citations de brevets; la date doit être indiquée conformément à la norme ST.2 de l'OMPI (CCYY MM DD)
<313>	Résidus pertinents dans SEQ ID NO: x	F	
<400>	Séquence	O	le SEQ ID NO : x doit suivre l'identificateur numérique et figurer sur la ligne précédant la séquence (voir l'appendice 3)

[L'appendice 2 de l'annexe C suit]





**Annexe C, appendice 2**  
**Symboles des nucléotides et des acides aminés**  
**et tableau des caractéristiques**

**Tableau 1 : liste des nucléotides**

Symbole	Signification	Origine de la désignation
a	a	<u>a</u> dénine
g	g	g <u>u</u> anine
c	c	<u>c</u> ytosine
t	t	<u>t</u> hymine
u	u	<u>u</u> racile
r	g ou a	pur <u>r</u> ine
y	t/u ou c	pyr <u>y</u> midine
m	a ou c	am <u>m</u> ino
k	g ou t/u	<u>k</u> eto (Ceto)
s	g ou c	interactions fortes (liaisons 3 H)
w	a ou t/u	interactions faibles (liaisons 2 H)
b	g ou c ou t/u	autre que a
d	a ou g ou t/u	autre que c
h	a ou c ou t/u	autre que g
v	a ou g ou c	autre que t et u
n	a, g, c t/u, non connu, ou autre	<u>n</u> 'importe lequel



[Annexe C, appendice 2, suite]

**Tableau 2 : liste des nucléotides modifiés**

<b>Symbole</b>	<b>Signification</b>
ac4c	4-acétylcytidine
chm5u	5-(carboxyhydroxyméthyl)uridine
cm	2'-O-méthylcytidine
cmnm5s2u	5-carboxyméthylaminométhyl-2-thiouridine
cmnm5u	5-carboxyméthylaminométhyluridine
d	dihydrouridine
fm	2'-O-méthylpseudouridine
gal q	bêta, D-galactosylquéosine
gm	2'-O-méthylguanosine
i	inosine
i6a	N6-isopentényladénosine
m1a	1-méthyladénosine
m1f	1-méthylpseudouridine
m1g	1-méthylguanosine
m1i	1-méthylinosine
m22g	2,2-diméthylguanosine
m2a	2-méthyladénosine
m2g	2-méthylguanosine
m3c	3-méthylcytidine
m5c	5-méthylcytidine
m6a	N6-méthyladénosine
m7g	7-méthylguanosine
mam5u	5-méthylaminométhyluridine
mam5s2u	5-méthoxyaminométhyl-2-thiouridine
man q	bêta, D-mannosylquéosine
mcm5s2u	5-méthoxycarbonylméthyl-2-thiouridine
mcm5u	5-méthoxycarbonylméthyluridine
mo5u	5-méthoxyuridine
ms2i6a	2-méthylthio-N6-isopentényladénosine
ms2t6a	N-((9-bêta-D-ribofuranosyl-2-méthylthiopurine-6-yl) carbamoyl) thréonine



MANUEL SUR L'INFORMATION ET LA DOCUMENTATION EN MATIÈRE DE PROPRIÉTÉ INDUSTRIELLE

Réf. : Normes – ST.25

page : 3.25.19

[Annexe C, appendice 2, suite]

mt6a	N-((9-bêta-D-ribofuranosylpurine-6-yl)N-méthylcarbamoyl) thréonine
mv	5-méthoxycarbonylméthoxyuridine
o5u	5-carboxyméthoxyuridine
osyw	wybutoxosine
p	pseudouridine
q	quéuosine
s2c	2-thiocytidine
s2t	5-méthyl-2-thiouridine
s2u	2-thiouridine
s4u	4-thiouridine
t	5-méthyluridine
t6a	N-((9-bêta-D-ribofuranosylpurine-6-yl)-carbamoyl) thréonine
tm	2'-O-méthyl-5-méthyluridine
um	2'-O-méthyluridine
yw	wybutosine
x	3-(3-amino-3-carboxypropyl)uridine, (acp3)u



[Annexe C, appendice 2, suite]

**Tableau 3 : liste des acides aminés**

<b>Symbole</b>	<b>Signification</b>
Ala	Alanine
Cys	Cystéine
Asp	Acide aspartique
Glu	Acide glutamique
Phe	Phénylalanine
Gly	Glycine
His	Histidine
Ile	Isoleucine
Lys	Lysine
Leu	Leucine
Met	Méthionine
Asn	Asparagine
Pro	Proline
Gln	Glutamine
Arg	Arginine
Ser	Sérine
Thr	Thréonine
Val	Valine
Trp	Tryptophane
Tyr	Tyrosine
Asx	Asp ou Asn
Glx	Glu ou Gln
Xaa	non connu ou autre



[Annexe C, appendice 2, suite]

**Tableau 4 : liste des acides aminés modifiés ou peu connus**

<b>Symbole</b>	<b>Signification</b>
Aad	acide 2-aminoadipique
bAad	acide 3-aminoadipique
bA1a	bêta-alanine, bêta-acide aminopropionique
Abu	acide 2-aminobutyrique
4Abu	acide 4-aminobutyrique, acide pipéridinique
Acp	acide 6-aminocaproïque
Ahe	acide 2-aminoheptanoïque
Aib	acide 2-aminoisobutyrique
bAib	acide 3-aminoisobutyrique
Apm	acide 2-aminopimélique
Dbu	acide 2,4-diaminobutyrique
Des	desmosine
Dpm	acide 2,2-diaminopimélique
Dpr	acide 2,3-diaminopropionique
EtGly	N-éthylglycine
EtAsn	N-éthylasparagine
Hyl	hydroxylysine
aHyl	allo-hydroxylysine
3Hyp	3-hydroxyproline
4Hyp	4-hydroxyproline
Ide	isodesmosine
alle	allo-isoleucine
MeGly	N-méthylglycine, sarcosine
MeIle	N-méthylisoleucine
MeLys	6-N-méthyllysine
MeVal	N-méthylvaline
Nva	norvaline
Nle	norleucine
Orn	ornithine



[Annexe C, appendice 2, suite]

**Tableau 5 : liste des clés de caractérisation concernant les séquences de nucléotides**

Clé	Description
allele	Individu ou souche apparenté contenant des formes stables différentes d'un même gène, qui se distingue de la séquence présentée à cet emplacement (et éventuellement à d'autres emplacements)
attenuator	1) région d'ADN où se produit une régulation de la terminaison de la transcription qui contrôle l'expression de certains opérons bactériens; 2) segment de séquence situé entre le promoteur et le premier gène de structure, qui provoque une terminaison partielle de la transcription
C_region	Région constante de la chaîne lourde et de la chaîne légère de l'immunoglobuline et des chaînes alpha, bêta et gamma du récepteur d'un lymphocyte T; comprend un ou plusieurs exons, selon la chaîne
CAAT_signal	Séquence CAAT; partie d'une séquence conservée située environ 75 paires de bases en amont du site d'initiation des unités de transcription eucaryotes, qui peut jouer un rôle dans la fixation de l'ARN polymérase; consensus=GG (C ou T) CAATCT
CDS	Séquence codante ("coding sequence"); séquence de nucléotides correspondant à celle des acides aminés dans une protéine (l'emplacement comprend le codon d'arrêt); contient la traduction conceptuelle des acides aminés
conflict	les déterminations indépendantes de la "même" séquence diffèrent sur ce site ou dans cette région
D-loop	Boucle de déplacement (" <u>displacement loop</u> "); région au sein de l'ADN mitochondrial où une petite partie d'ARN est appariée à un brin d'ADN, entraînant le déplacement du brin original d'ADN dans cette région; désigne aussi le déplacement d'une région d'ADN double brin sous l'effet d'un envahisseur simple brin dans la réaction catalysée par la protéine RecA
D-segment	Segment de diversité (" <u>diversity segment</u> ") de la chaîne lourde de l'immunoglobuline et de la chaîne bêta du récepteur d'un lymphocyte T
enhancer	Séquence en cis entraînant l'utilisation accrue de (certains) promoteurs eucaryotes et dont l'action s'exerce quelle que soit l'orientation et l'emplacement (en amont ou en aval) par rapport au promoteur
exon	Région du génome codant une partie de l'ARN messager épissé; peut contenir la région 5'UTR, tous les CDS et la région 3'UTR
GC_signal	Séquence GC; région conservée riche en GC, située en amont du site d'initiation des unités de transcription eucaryotes, qui peut prendre la forme de copies multiples et se produire dans les deux sens; consensus=GGGCGG
gene	Région présentant un intérêt biologique, identifiée comme étant un gène et à laquelle un nom a été attribué
iDNA	AND intercalaire; ADN éliminé par l'un des types de recombinaison.
intron	Segment d'ADN qui est transcrit, puis éliminé par aboutement des séquences (exons) situées de part et d'autre



## MANUEL SUR L'INFORMATION ET LA DOCUMENTATION EN MATIÈRE DE PROPRIÉTÉ INDUSTRIELLE

Réf. : Normes – ST.25

page : 3.25.23

[Annexe C, appendice 2, suite]

Clé	Description
J_segment	Segment de jonction (“ <u>joining segment</u> ”) de la chaîne légère et de la chaîne lourde de l’immunoglobuline, ainsi que des chaînes alpha, bêta et gamma du récepteur d’un lymphocyte T
LTR	Répétition terminale longue (“ <u>long terminal repeat</u> ”); séquence directement répétée aux deux extrémités d’une séquence définie, du type de celle que l’on trouve dans les rétrovirus
mat_peptide	Séquence codante d’un peptide mature ou d’une protéine mature; séquence codante du peptide ou de la protéine à l’état mature ou final, qui suit la modification post-traductionnelle; l’emplacement ne comprend pas le codon d’arrêt (contrairement au CDS correspondant)
misc_binding	site d’un acide nucléique fixant, par covalence ou non, un autre fragment de molécule, qui ne peut être décrit par aucune autre clé de fixation (primer_bind ou protein_bind)
misc_difference	séquence de caractérisation différente de celle qui est présentée dans l’entrée et ne pouvant pas être décrite par une autre clé de différence (conflict, unsure, old_sequence, mutation, variation, allele ou modified_base)
misc_feature	région présentant un intérêt biologique, qui ne peut pas être décrite par une autre clé de caractérisation; nouvelle caractéristique ou caractéristique rare
misc_recomb	site de toute recombinaison généralisée, spécifique d’un site ou répllicative, où se produit la cassure et la réunion de l’ADN double brin et qui ne peut pas être décrite par une autre clé de recombinaison (iDNA ou virion) ou par un autre qualificateur de clé source (/insertion_seq, /transposon, /proviral)
misc_RNA	tout transcrit ou produit de l’ARN qui ne peut pas être défini par une autre clé de l’ARN (prim_transcript, precursor_RNA, mRNA, 5’clip, 3’clip, 5’UTR, 3’UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, rRNA, tRNA, scRNA ou snRNA)
misc_signal	toute région contenant un signal qui commande ou modifie une fonction ou l’expression d’un gène, qui ne peut pas être décrite par une autre clé de signal (promoter, CAAT_signal, TATA_signal, –35_signal, –10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator ou rep_origin)
misc_structure	toute structure ou conformation secondaire ou tertiaire qui ne peut pas être décrite par une autre clé de structure (stem_loop ou D-loop)
modified_base	le nucléotide indiqué est un nucléotide modifié, qui doit être remplacé par la molécule indiquée (donnée dans la valeur qualificative mod_base)
mRNA	ARN messenger; comprend la région non traduite en 5’ (5’UTR), les séquences codantes (CDS, exon) et la région non traduite en 3’ (3’UTR)
mutation	la souche apparentée présente un changement brusque et transmissible dans la séquence, à cet emplacement
N_region	des nucléotides supplémentaires sont insérés entre des segments d’immunoglobuline réarrangés
old_sequence	la séquence présentée est la version modifiée d’une ancienne séquence à cet emplacement



## MANUEL SUR L'INFORMATION ET LA DOCUMENTATION EN MATIÈRE DE PROPRIÉTÉ INDUSTRIELLE

Réf. : Normes – ST.25

page : 3.25.24

[Annexe C, appendice 2, suite]

Clé	Description
polyA_signal	site de reconnaissance indispensable à la coupure d'un transcrit par endonucléase, suivie d'une polyadénylation; consensus=AATAAA
polyA_site	site d'un transcrit auquel sont ajoutés des résidus d'adénine par polyadénylation post-transcriptionnelle
precursor_RNA	ARN précurseur, c'est-à-dire tout type d'ARN qui n'est pas encore mature; peut comprendre la région coupée en 5' (5'clip), la région non traduite en 5' (5'UTR), les séquences codantes (CDS, exon), les séquences intercalaires (intron), la région non traduite en 3' (3'UTR) et la région coupée en 3' (3'clip)
prim_transcript	transcrit primaire (initial, non remanié); comprend la région coupée en 5' (5'clip), la région non traduite en 5' (5'UTR), les séquences codantes (CDS, exon), les séquences intercalaires (intron), la région non traduite en 3' (3'UTR) et la région coupée en 3' (3'clip)
primer_bind	site de fixation non covalent pour amorces dans l'initiation de la réplication, de la transcription ou de la transcription inverse; comprend les sites pour les éléments de synthèse, par exemple les amorces de l'amplification en chaîne par polymérase (PCR)
promoter	région d'une molécule d'ADN jouant un rôle dans la fixation de l'ARN polymérase en vue de l'initiation de la transcription
protein_bind	site de fixation non covalent des protéines sur un acide nucléique
RBS	site de fixation du ribosome ("ribosome binding site")
repeat_region	région du génome contenant des unités de répétition
repeat_unit	unité d'un élément de répétition
rep_origin	origine de la réplication; site de départ pour la duplication d'un acide nucléique en vue de l'obtention de deux copies identiques
rRNA	ARN ribosomique mature; molécule d'ARN de la particule ribonucléoprotéique (ribosome) qui assemble les acides aminés en protéines
S_region	région de commutation ("switch region") des chaînes lourdes de l'immunoglobuline; joue un rôle dans le réarrangement de la chaîne lourde de l'ADN, qui conduit à l'expression d'une classe d'immunoglobuline différente à partir du même lymphocyte B
satellite	nombreuses séquences répétées en tandem (identiques ou apparentées) d'une courte unité de répétition de base; nombre d'entre elles ont une composition de base ou une propriété différente de la moyenne du génome, qui leur permet d'être séparées du reste de l'ADN génomique (bande principale)
scRNA	petit ARN cytoplasmique ("small cytoplasmic RNA"); l'une des nombreuses petites molécules d'ARN cytoplasmique présentes dans le cytoplasme et (parfois) dans le noyau d'un eucaryote
sig_peptide	séquence codante d'un peptide-signal; séquence codante d'un N-terminal de protéine sécrétée; ce domaine joue un rôle dans l'intégration du polypeptide naissant dans la membrane; séquence leader





## MANUEL SUR L'INFORMATION ET LA DOCUMENTATION EN MATIÈRE DE PROPRIÉTÉ INDUSTRIELLE

Réf. : Normes – ST.25

page : 3.25.25

[Annexe C, appendice 2, suite]

Clé	Description
snRNA	petit ARN nucléaire (“small nuclear RNA”); l’une des nombreuses petites espèces d’ARN confinées au noyau; plusieurs snRNA jouent un rôle dans l’excision-épissage ou dans d’autres réactions de maturation moléculaire de l’ARN
source	permet d’identifier la source biologique de l’intervalle de séquence indiqué; cette clé est obligatoire; chaque entrée doit comporter, au minimum, une clé source unique couvrant la séquence tout entière; il est possible d’utiliser plus d’une clé source par séquence
stem_loop	épingle à cheveux; région d’une double hélice formée par l’appariement de bases entre des séquences contiguës (inversées) complémentaires appartenant à un même brin d’ARN ou d’ADN
STS	site de séquence étiqueté; séquence d’ADN courte et unique caractérisant un point de repère de la cartographie du génome et qui peut être détectée moyennant une amplification en chaîne par polymérase (PCR); la carte d’une région du génome peut être dressée par détermination de l’ordre d’une série de STS
TATA_signal	séquence TATA; séquence de Goldberg-Hogness; heptamère conservé, riche en A et T, situé environ 25 paires de bases en amont du site d’initiation de chaque unité transcrite par l’ARN polymérase II des eucaryotes, qui peut jouer un rôle dans le positionnement de l’enzyme aux fins d’une initiation correcte; consensus=TATA(A ou T)A(A ou T)
terminator	séquence d’ADN située soit à l’extrémité du transcrit, soit à côté d’un promoteur qui conduit l’ARN polymérase à terminer la transcription; peut aussi être le site de fixation d’un répresseur
transit_peptide	séquence codante d’un peptide-transit; séquence codante d’un N-terminal de protéine d’un organite codée par le noyau; ce domaine joue un rôle dans l’importation post-traductionnelle de la protéine dans l’organite
tRNA	ARN de transfert mature; courte molécule d’ARN (75-85 bases) qui permet la traduction d’une séquence d’acides nucléiques en une séquence d’acides aminés
unsure	l’auteur n’est pas certain de l’exactitude de la séquence dans cette région
V_region	région variable de la chaîne légère et de la chaîne lourde de l’immunoglobuline et des chaînes alpha, bêta et gamma du récepteur d’un lymphocyte T; codes applicables à la portion variable du terminal amino; peut être composée des segments suivants : V_segments, D_segments, N_régions et J_segments
V_segment	segment variable de la chaîne légère et de la chaîne lourde de l’immunoglobuline et des chaînes alpha, bêta et gamma du récepteur d’un lymphocyte T; codes applicables à la plus grande partie de la région variable (V_région) et aux quelques acides aminés du peptide leader qui subsistent
variation	souche apparentée contenant des mutations stables du même gène (par exemple : RFLP, polymorphismes, etc.) qui diffèrent de la séquence présentée à cet emplacement (et éventuellement à d’autres emplacements)
3’clip	région d’un transcrit précurseur située en position 3’, qui est coupée durant la maturation moléculaire



[Annexe C, appendice 2, suite]

Clé	Description
3'UTR	région en position 3' à l'extrémité d'un transcrit mature (qui suit le codon de terminaison) qui n'est pas traduite en protéine
5'clip	région en position 5' d'un transcrit précurseur qui est coupée durant la maturation moléculaire
5'UTR	région en position 5' située à l'extrémité d'un transcrit mature (avant le codon d'initiation) qui n'est pas traduite en protéine
-10_signal	séquence de pribnow; région conservée située à environ 10 paires de bases en amont du site d'initiation des unités de transcription bactériennes, qui peut jouer un rôle dans la fixation de l'ARN polymérase; consensus=TAtAaT
-35_signal	hexamère conservé situé à environ 35 paires de bases en amont du site d'initiation des unités de transcription bactériennes; consensus=TTGACa [ ] ou TGTTGACA [ ]



[Annexe C, appendice 2, suite]

**Tableau 6 : liste des clés de caractérisation concernant les séquences de protéines**

Clé	Description
CONFLICT	séquences différentes selon divers documents
VARIANT	les auteurs signalent qu'il existe des variants de la séquence
VARSPLIC	description des variants de la séquence produits par une excision-épissage différentielle
MUTAGEN	site modifié à titre expérimental
MOD_RES	modification post-traductionnelle d'un résidu
ACETYLATION	N-terminal ou autre
AMIDATION	en général, au C-terminal d'un peptide mature actif
BLOCKED	groupe de blocage N- ou C-terminal indéterminé
FORMYLATION	de la méthionine N-terminale
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	de l'asparagine, de l'acide aspartique, de la proline ou de la lysine
METHYLATION	en général, de la lysine ou de l'arginine
PHOSPHORYLATION	de la sérine, de la thréonine, de la tyrosine, de l'acide aspartique ou de l'histidine
PYRROLIDONE CARBOXYLIC ACID	glutamate N-terminal ayant formé un lactame cyclique interne
SULFATATION	en général, de la tyrosine
LIPID	liaison covalente d'un fragment lipidique
MYRISTATE	groupe myristate rattaché, par une liaison amide, au résidu N-terminal glycine de la forme mature d'une protéine ou à un résidu de lysine interne
PALMITATE	groupe palmitate lié, par une liaison thioether, à un résidu de cystéine ou, par une liaison ester, à un résidu de sérine ou de thréonine
FARNESYL	groupe farnésol rattaché, par une liaison thioether, à un résidu de cystéine
GERANYL-GERANYL	groupe géranyl-géranyl rattaché, par une liaison thioether, à un résidu de cystéine
GPI-ANCHOR	groupe glycosyl-phosphatidylinositol (GPI) lié au groupe alpha carboxyl du résidu C-terminal de la forme mature d'une protéine
N-ACYL DIGLYCERIDE	cystéine N-terminale de la forme mature d'une lipoprotéine procaryote assortie d'un acide gras amidé et d'un groupe glycéride auquel deux acides gras sont rattachés par des liaisons ester



## MANUEL SUR L'INFORMATION ET LA DOCUMENTATION EN MATIÈRE DE PROPRIÉTÉ INDUSTRIELLE

Réf. : Normes – ST.25

page : 3.25.28

[Annexe C, appendice 2, suite]

Clé	Description
DISULFID	liaison disulfure; les extrémités "FROM" et "TO" représentent les deux résidus qui sont liés par une liaison disulfure intrachaîne; si les extrémités "FROM" à "TO" sont identiques, la liaison disulfure est une liaison interchaîne et le champ de description donne la nature de la liaison réticulée
THIOLEST	liaison thiolester; les extrémités "FROM" et "TO" représentent les deux résidus liés par la liaison thiolester
THIOETH	liaison thio-éther; les extrémités "FROM" et "TO" représentent les deux résidus liés par la liaison thio-éther
CARBOHYD	site de glycosylation; la nature de l'hydrate de carbone (lorsqu'elle est connue) est donnée dans le champ de description
METAL	site de fixation pour un ion métallique; le champ de description donne la nature du métal
BINDING	site de fixation pour tout groupe chimique (coenzyme, groupe prosthétique, etc.); la nature chimique du groupe est donnée dans le champ de description
SIGNAL	séquence-signal (prépeptide)
TRANSIT	peptide-transit (mitochondrial, chloroplastique ou destiné à un micro-organisme)
PROPEP	propeptide
CHAIN	chaîne polypeptidique dans la protéine mature
PEPTIDE	peptide actif libéré
DOMAIN	domaine d'intérêt dans la séquence; la nature de ce domaine est donnée dans le champ de description
CA_BIND	région de fixation du calcium
DNA_BIND	région de fixation de l'ADN
NP_BIND	région de fixation d'un phosphate nucléotidique; la nature du phosphate nucléotidique est donnée dans le champ de description
TRANSMEM	région transmembranaire
ZN_FING	région d'un doigt de zinc
SIMILAR	similitude avec une autre séquence protéique; des informations détaillées sur cette séquence figurent dans le champ de description
REPEAT	répétition de séquence interne
HELIX	structure secondaire – Hélices, telles que les hélices alpha, les hélices 3-10 ou les hélices pi
STRAND	structure secondaire – Brin bêta, tel que le brin bêta à liaison hydrogénée ou le résidu dans un brin isolé à pont bêta
TURN	structure secondaire – Virages, tels que le virage à liaison H (virage 3, virage 4 ou virage 5)



[Annexe C, appendice 2, suite]

Clé	Description
ACT_SITE	acides aminés jouant un rôle dans l'activité de l'enzyme
SITE	tout autre site présentant un intérêt dans la séquence
INIT_MET	la séquence commence par un initiateur méthionine
NON_TER	le résidu situé à une extrémité de la séquence n'est pas le résidu terminal; appliqué à la position 1, cela signifie que la première position n'est pas la position N-terminale de la molécule complète; s'il est appliqué à la dernière position, cela signifie que cette position n'est pas la position C-terminale de la molécule complète; il n'y a pas de champ de description pour cette clé
NON_CONS	résidus non consécutifs; indique que deux résidus dans une séquence ne sont pas consécutifs et qu'il existe un certain nombre de résidus non séquencés entre eux
UNSURE	zones d'incertitude dans la séquence; sert à décrire les régions d'une séquence pour lesquelles les auteurs ne sont pas sûrs de la définition

[L'appendice 3 de l'annexe C suit]



Annexe C, appendice 3
Exemple de listage des séquences

<110> Smith, John; Smithgene Inc.
<120> Example of a Sequence Listing
<130> 01-00001
<140> PCT/EP98/00001
<141> 1998-12-31
<150> US 08/999,999
<151> 1997-10-15
<160> 4
<170> PatentIn version 2.0
<210> 1
<211> 389
<212> DNA
<213> Paramecium sp.
<220>
<221> CDS
<222> (279)...(389)
<300>
<301> Doe, Richard
<302> Isolation and Characterization of a Gene Encoding a Protease from Paramecium sp.
<303> Journal of Genes
<304> 1
<305> 4
<306> 1-7
<307> 1988-06-31
<308> 123456
<309> 1988-06-31
<400> 1
agctgtagtc attcctgtgt cctcttctct ctgggcttct Caccctgcta atcagatctc 60
agggagagtg tcttgaccct cctctgcctt tgcagcttca Caggcaggca ggcaggcagc 120
tgatgtggca attgctggca gtgccacagg cttttcagcc Aggcttaggg tgggttccgc 180
cgcggcgcgg cggcccctct cgcgctctc tcgcgcctct Ctctcgtctt cctctcgtctc 240
ggacctgatt aggtgagcag gaggaggggg cagttagc atg Gtt tca atg ttc agc 296
Met Val Ser Met Phe Ser
1 5
ttg tct ttc aaa tgg cct gga ttt tgt ttg ttt Gtt tgt ttg ttc caa 344
Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu Phe Val Cys Leu Phe Gln
10 15 20
tgt ccc aaa gtc ctc ccc tgt cac tca tca ctg Cag ccg aat ctt 389
Cys Pro Lys Val Leu Pro Cys His Ser Ser Leu Gln Pro Asn Leu
25 30 35



## MANUEL SUR L'INFORMATION ET LA DOCUMENTATION EN MATIÈRE DE PROPRIÉTÉ INDUSTRIELLE

Réf. : Normes – ST.25

page : 3.25.31

[Annexe C, appendice 3, suite]

<210> 2  
<211> 37  
<212> PRT  
<213> Paramecium sp.

<400> 2  
Met Val Ser Met Phe Ser Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu  
1 5 10 15  
Phe Val Cys Leu Phe Gln Cys Pro Lys Val Leu Pro Cys His Ser Ser  
20 25 30  
Leu Gln Pro Asn Leu  
35

<210> 3  
<211> 11  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Designed peptide based on size and polarity to act as a linker  
between the alpha and beta chains of Protein XYZ.

<400> 3  
Met Val Asn Leu Glu Pro Met His Thr Glu Ile  
1 5 10

<210> 4  
<400> 4  
000

[Fin de l'appendice 3 et de la norme]