



## Annexe 1

[relatif à l'art. 11.1), deuxième phrase, Ordonnance relative aux brevets]

Normes relatives au dépôt de listages des séquences

**Sans validité à compter du 1er juillet 2022**

**Dienststelle München**

**Dienststelle Jena**

**Informations- und Dienstleistungszentrum Berlin**

**Adresse postale**

80297 München

07738 Jena

10958 Berlin

**Télécopieur**

+49 89 2195-2221

+49 3641 40-5690

+49 30 25992-404

**Téléphone**

**Service central à la clientèle :**

**+49 89 2195-1000**

**Bénéficiaire :** Bundeskasse Halle/DPMA

IBAN : DE84 7000 0000 0070 0010 54, BIC (SWIFT-Code) : MARKDEF1700

**Adresse de la banque :** Bundesbankfiliale München, Leopoldstr. 234, 80807 München

**Internet :**

<https://www.dpma.de>

## Contenu

Définitions.....	3
Listage des séquences.....	3
Séquences de nucléotides .....	4
Symboles à utiliser .....	4
Mode de présentation à suivre.....	4
Séquences d'acides aminés.....	4
Symboles à utiliser .....	4
Mode de présentation à suivre.....	4
Autres renseignements devant figurer dans le listage des séquences.....	5
Éléments de données obligatoires .....	5
Éléments de données facultatifs .....	6
Présentation des caractéristiques .....	6
Texte libre .....	6
Listage des séquences déposé ultérieurement .....	6
Listage des séquences sous une forme déchiffrable par ordinateur .....	6
Identificateurs numériques admissibles.....	7
Tableau 1 Liste des nucléotides.....	10
Tableau 2 Liste des nucléotides modifiés.....	10
Tableau 3 Liste des acides aminés.....	11
Tableau 4 Liste des acides aminés modifiés ou peu connus.....	12
Tableau 5 Liste des clés de caractérisation concernant les séquences de nucléotides.....	12
Tableau 6 Liste des clés de caractérisation concernant les séquences de protéines .....	16
Exemple .....	18

Sans validité à compter du 1er juillet 2022

## Définitions

1. Aux fins de la présente norme,
  - i) l'expression « listage des séquences » désigne une partie de la description figurant dans la demande au moment de son dépôt, ou un document déposé après la demande, qui divulgue de façon détaillée les séquences de nucléotides ou d'acides aminés ainsi que d'autres informations disponibles ;
  - 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 du paragraphe 48 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 du paragraphe 48. Les modifications (par exemple, les bases méthylées) peuvent être décrites de la manière indiquée dans le tableau 2 du paragraphe 48, 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 du paragraphe 48. 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 du paragraphe 48, les positions modifiées (par exemple hydroxylations ou glycosylations) étant elles-mêmes décrites de la manière indiquée dans le tableau 4 du paragraphe 48 ; 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 du paragraphe 48 et accompagné(e), par exemple, d'une description des liaisons anormales, des liaisons croisées (par exemple 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 du paragraphe 48, et les clés de caractérisation, dans les tableaux 5 et 6 du paragraphe 48) ;

- viii) l'expression « administration compétente » désigne l'administration chargée de la recherche internationale ou l'administration chargée de l'examen préliminaire international pour la demande internationale en question, ou encore l'office désigné ou élu au sein duquel le traitement de la demande internationale a commencé.

## Listage des séquences

2. Le listage des séquences défini au paragraphe 1.i) doit, lorsqu'il est déposé en même temps que la demande, être placé à la fin de celle-ci. Cette partie de la demande doit s'intituler « *Sequenzprotokoll* » ou « *Sequence Listing* » (listage des séquences), commencer de préférence sur une nouvelle page et faire l'objet d'une pagination distincte. Le listage des séquences fait partie intégrante de la description ; il n'est donc pas nécessaire, sous réserve du paragraphe 35, de décrire les séquences ailleurs dans la description.
3. Lorsque le listage des séquences défini au paragraphe 1.i) ne figure pas dans la demande telle qu'elle a été déposée mais constitue un document distinct remis après le dépôt de la demande (voir le paragraphe 36), il doit s'intituler « *Sequenzprotokoll* » ou « *Sequence Listing* » (listage des séquences) et faire l'objet d'une pagination distincte. La numérotation des séquences (voir le paragraphe 4) figurant dans la demande telle qu'elle a été déposée doit être maintenue dans un listage des séquences remis ultérieurement.
4. 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.
5. 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 : ».
6. 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

7. 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.
8. 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 du paragraphe 48 doivent être utilisées.
9. 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 du paragraphe 48 ; 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 du paragraphe 48. 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 31). Le symbole « n » est l'équivalent d'un seul nucléotide inconnu ou modifié.

### Mode de présentation à suivre

10. Une séquence de nucléotides doit comporter 60 bases par ligne au maximum, avec un espace entre chaque codon ou groupe de dix bases.
11. 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 dix) qui restent à l'extrémité des régions non codantes d'une séquence doivent être regroupées et séparées des groupes voisins par un espace.
12. Les bases des régions codantes d'une séquence de nucléotides doivent figurer sur la liste sous forme de triplets (codons).
13. 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.
14. 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

15. 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.
16. 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 du paragraphe 48. 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 21).
17. 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 du paragraphe 48 ; 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 du paragraphe 48. 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 31). Le symbole « Xaa » est l'équivalent d'un seul acide aminé inconnu ou modifié.

### Mode de présentation à suivre

18. Une séquence de protéines ou de peptides doit comporter seize acides aminés par ligne au maximum, avec un espace entre chaque acide aminé.
19. 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.
20. 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.
21. 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 renseignements devant figurer dans le listage des séquences

22. 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 au paragraphe 47.
23. Seuls les identificateurs numériques des éléments de données définis au paragraphe 47 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 les 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

24. 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 au paragraphe 47 (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 4 et la séquence SEQ ID NO : 4 dans l'exemple donné à la fin de la présente norme).

25. Outre les éléments de données indiqués au paragraphe 24, lorsqu'un listage des séquences est déposé en même temps que la demande à laquelle il se rapporte ou 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>	Numéro de référence
-------	---------------------

26. Outre les éléments de données énumérés au paragraphe 24, lorsqu'un listage des séquences est déposé sur requête d'une administration compétente ou 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

27. Outre les éléments de données indiqués au paragraphe 24, 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

28. 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

29. 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

30. Tous les éléments de données définis au paragraphe 47 qui ne sont pas mentionnés dans les paragraphes 24 à 29 sont facultatifs (éléments de données facultatifs).

### Présentation des caractéristiques

31. Les caractéristiques des séquences (identificateur numérique <220>) doivent être décrites à l'aide des « clés de caractérisation » indiquées dans les tableaux 5 et 6<sup>1</sup> du paragraphe 48.

### Texte libre

32. 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 1.vii).
33. 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. Toute autre information doit figurer dans la partie principale de la description dans la langue de celle-ci.
34. Le texte libre peut être rédigé en allemand ou en anglais.
35. Lorsque, dans la description, la partie réservée au listage des séquences 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 partie principale de la description soit inséré dans une rubrique particulière de la description intitulée « *Sequenzprotokoll - freier Text* » ou « *Sequence Listing Free Text* » (texte libre du listage des séquences).

### Listage des séquences déposé ultérieurement

36. Tout listage des séquences qui ne figure pas dans la demande telle qu'elle a été déposée et qui est remis ultérieurement ne doit pas contenir d'éléments allant au-delà de la divulgation faite dans ladite demande et doit être accompagné d'une déclaration à cet effet. Autrement dit, un listage des séquences remis après le dépôt de la demande ne doit contenir que les séquences qui ont été divulguées dans la demande en question.
37. Des listages des séquences ne figurant pas dans la demande telle qu'elle a été déposée ne font pas partie de la divulgation de l'invention. En vertu de l'article 11.3), il est possible qu'un listage des séquences figurant dans la demande telle qu'elle a été déposée fasse l'objet d'une correction.

### Listage des séquences sous une forme déchiffrable par ordinateur

38. Outre le listage des séquences figurant dans la demande, une copie de ce même listage doit être fournie sous une forme déchiffrable par ordinateur.
39. Tout listage des séquences sous une forme déchiffrable par ordinateur qui est remis en sus du listage des séquences présenté par écrit doit être identique à ce dernier et être accompagné de la déclaration suivante : « les informations enregistrées sous une forme déchiffrable par ordinateur sont identiques à celles du listage des séquences présenté par écrit ».
40. La copie imprimable du listage des séquences doit, de préférence, figurer tout entière dans un seul fichier électronique sur une seule disquette ou sur tout autre support électronique admis par l'Office allemand des brevets et des marques. Le fichier enregistré sur la disquette ou sur tout autre support électronique admis par l'administration compétente doit être codé selon la page de code IBM<sup>2</sup> 437, la page de code IBM 932<sup>3</sup> ou une page de code compatible. Une page de code compatible (requise 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. Les types de supports et les formatages suivants sont acceptés pour des listages de séquences fournis également sous forme déchiffrable par machine :

Supports	Type	Formatage
CD-R	120 mm Recordable Disk	ISO 9660
DVD-R	120 mm DVD-Recordable Disk (4,7 GB)	conforme à ISO 9660 ou OSTA UDF (1.02 ou plus élevé)
DVD+R	120 mm DVD-Recordable Disk (4,7 GB)	conforme à ISO 9660 ou OSTA UDF (1.02 ou plus élevé)

42. Le listage déchiffrable par ordinateur peut être créé par tout moyen. Toutefois, il doit répondre au formats spécifiés par l'Office allemand des brevets et des marques. Il devrait être créé de préférence par un logiciel spécialisé tel que PatentIn.
43. En cas d'utilisation d'un support de données matériel, la compression d'un fichier est admise dans la mesure où le fichier compressé se présente sous un format auto-extractible qui se décompressera sur un système d'exploitation (MS Windows) spécifié par l'Office allemand des brevets et des marques. De même, des fichiers ayant un contenu connexe peuvent être compressés dans un format non auto-extractible, pourvu que le fichier d'archive est présenté au format

<sup>1</sup> 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).

<sup>2</sup> IBM est une marque enregistrée de la société International Business Machines, des Etats-Unis d'Amérique.

<sup>3</sup> Les pages de code mentionnées constituent des normes *de facto* pour les ordinateurs personnels.

ZIP dans la version du 13 juillet 1998 et qu'il ne contient pas d'autres archives ZIP ni de structure de répertoire.

44. Sur le support de données matériel, il faut apposer une étiquette fixe portant les indications manuscrites en majuscules d'imprimerie ou dactylographiées suivantes: le nom du déposant, le titre de l'invention, un numéro de référence, la date à laquelle les données ont été enregistrées et le système d'exploitation informatique.
45. Si le support de données matériel est fourni après la date de dépôt de la demande, cette date et le numéro de la demande doivent aussi figurer sur les étiquettes. Des corrections ou autres modifications relatives au listage des séquences doivent être présentées aussi bien par écrit que sous forme déchiffrable par ordinateur.
46. Toute correction de la version imprimée du listage des séquences, présentée en vertu des règles 13<sup>ter</sup> 1.a)i) ou 26.3 du règlement d'exécution du PCT, toute rectification d'une erreur évidente de la version imprimée, présentée en vertu de la règle 91 du règlement d'exécution du PCT, et tout élément ajouté à la version imprimée du listage des séquences en

vertu de l'article 34 du PCT doit, en plus, être produit dans une version rectifiée, contenant les éléments additionnels, du listage des séquences sous forme déchiffrable par ordinateur.

#### 47. 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 24 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 25, 26, 27, 28 et 29 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 30 de la présente norme) sont signalés par la lettre « F ».

Identificateurs numériques admissibles			
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>	Numéro de référence	O, dans les cas mentionnés au paragraphe 25 de la présente norme	voir le paragraphe 25 de la présente norme
<140>	Demande de brevet actuelle	O, dans les cas mentionnés au paragraphe 26 de la présente norme	voir le paragraphe 26 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
<141>	Date de dépôt de la demande actuelle	O, dans les cas mentionnés au paragraphe 26 de la présente norme	voir le paragraphe 26 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 27 de la présente norme	voir le paragraphe 27 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 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)

Identificateurs numériques admissibles			
Identificateur numérique	Description de l'identificateur numérique	Obligatoire (O) ou facultatif (F)	Observations
<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 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 « inconnu »
<220>	Caractéristique	O, dans les cas mentionnés aux paragraphes 28 et 29 de la présente norme	laisser en blanc ; voir les paragraphes 28 et 29 de la présente norme ; description des motifs 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 28 de la présente norme	voir le paragraphe 28 de la présente norme ; seules doivent être utilisées les clés décrites dans les tableaux 5 ou 6 du paragraphe 48
<222>	Emplacement	O, dans les cas mentionnés au paragraphe 28 de la présente norme	voir le paragraphe 28 de la présente norme ; <ul style="list-style-type: none"> <li>- de (numéro de la première base ou du premier acide aminé dans la caractéristique)</li> <li>- à (numéro de la dernière base ou du dernier acide aminé dans la caractéristique)</li> <li>- bases (numéros des positions des bases dans une séquence de nucléotides)</li> <li>- acides aminés (numéros des positions des résidus d'acides aminés dans une séquence d'acides aminés)</li> <li>- lorsque la caractéristique se situe sur un brin complémentaire de celui figurant dans le listage des séquences</li> </ul>
<223>	Autres informations	O, dans les cas mentionnés aux paragraphes 28 et 29 de la présente norme	voir les paragraphes 28 et 29 de la présente norme ; toute autre information pertinente, exprimée à l'aide du vocabulaire non connoté ou sous forme de texte libre (en allemand ou en anglais) ; tout texte libre doit être répété dans la partie principale de la description, dans la même langue (voir le paragraphe 35 de la présente norme) ; lorsqu'une base modifiée ou un acide aminé modifié ou peu connu figurant dans les tableaux 2 et 4 du paragraphe 48 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 du paragraphe 48
<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

Identificateurs numériques admissibles			
Identificateur numérique	Description de l'identificateur numérique	Obligatoire (O) ou facultatif (F)	Observations
<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)
<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		le SEQ ID NO : x doit suivre l'identificateur numérique et figurer sur la ligne précédant la séquence (voir l'exemple)

Sans validité à compter du 1<sup>er</sup> juillet 2012

## 48. Symboles des nucléotides et des acides aminés et tableau des caractéristiques

Symbole	Signification	Origine de la désignation
a	a	<u>a</u> dénine
g	g	<u>g</u> uanine
c	c	<u>c</u> ytosine
t	t	<u>t</u> hymine
u	u	<u>u</u> racile
r	g ou a	<u>p</u> urine
y	t/u ou c	<u>p</u> yrimidine
m	a ou c	<u>a</u> mino
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 ou t/u, non connu, ou autre	<u>n</u> 'importe lequel (e any)

Symbole	Signification
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-galactosylquéuosine
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

Sans validité à compter du 1er juillet 2022

Symbole	Signification
m5c	5-méthylcytidine
m6a	N6-méthyladénosine
m7g	7-méthylguanosine
mam5u	5-méthylaminométhyluridine
mam5s2u	5-méthoxyaminométhyl-2-thiouridine
man q	bêta,D-mannosylquéuosine
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
mt6a	N-((9-bêta-D-ribofuranosylpurine-6-yl)N-méthylcarbamoyl)thréonine
mv	5-méthoxycarbonylméthoxyuridine
o5u	5-carboxyméthoxyuridine(v)
osyw	wybutosine
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

Symbole	Signification
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

Symbole	Signification
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

Symbole	Signification
EtAsn	N-éthylasparagine
Hyl	hydroxylysine
aHyl	allo-hydroxylysine
3Hyp	3-hydroxyproline
4Hyp	4-hydroxyproline
Ide	isodesmosine
alle	allo-isoleucine
MeGly	N-méthylglycine, sarcosine
Melle	N-méthylisoleucine
MeLys	6-N-méthyllysine
MeVal	N-méthylvaline
Nva	norvaline
Nle	norleucine
Orn	ornithine

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 prématurée 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 ( <u>coding sequence</u> ) ; 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 acide nucléique 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

Tableau 5 Liste des clés de caractérisation concernant les séquences de nucléotides	
Clé	Description
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, acide nucléique codant
iDNA	ADN 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
J_segment	segment de jonction (joining <u>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	séquence longue, 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 réplivative, 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 la 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)

Tableau 5  
Liste des clés de caractérisation concernant les séquences de nucléotides

Clé	Description
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
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 ( <u>ri</u> bosome <u>b</u> inding <u>s</u> ite)
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 l'ADN de chaînes lourdes, 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 ( <u>s</u> mall <u>c</u> ytoplasmic <u>R</u> NA) ; l'une des nombreuses petites olécules 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 rotéine sécrétée ; ce domaine joue un rôle dans l'intégration du polypeptide naissant dans la membrane ; séquence leader
snRNA	petit ARN nucléaire ( <u>s</u> mall <u>n</u> uclear <u>R</u> NA) ; 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

Tableau 5 Liste des clés de caractérisation concernant les séquences de nucléotides	
Clé	Description
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 rotéine 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
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

Tableau 6  
Liste des clés de caractérisation concernant les séquences de protéines

Clé	Description
CONFLICT	séquences différentes selon divers documents
VARIANT	les autres 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 thio-éther, 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 thio-éther, à un résidu de cystéine
GERANYL-GERANYL	groupe géranyl-géranyl rattaché, par une liaison thio-éther, 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
DISULFID	liaison disulfure ; les extrémités « FROM » et « TO » représentent les deux résidus qui sont liés par une liaison disulfure intrachaine ; 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

Tableau 6  
Liste des clés de caractérisation concernant les séquences de protéines

Clé	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)
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

**Exemple**

<110> Smith, John ; Smithgene Inc.  
 <120> Exemple de listage des séquences  
 <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 acagatctc 60  
 agggagagtg tcttgacct cctctgcctt tgcagcttca caggcaggca ggcaggcagc 120  
 tgatgtggca attgctggca gtgccacagg ctttcagcc agccttaggg tgggtccgc 180  
 cgcggcgcgg cggccccctc cgcgctctc tgcgctct ctctcgctct cctctcgctc 240  
 ggacctgatt aggtgagcag gaggaggggg cactagc 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 atc 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

Sans validité à compter du 1er juillet 2022