

## Beschluß des Präsidenten des Europäischen Patentamts vom 2. Oktober 1998 über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen

Der Präsident des Europäischen Patentamts, gestützt auf Artikel 10 EPÜ, Regeln 27a und 49 EPÜ sowie Regel 13<sup>ter</sup> PCT, beschließt:

### Artikel 1

*Standardisierte Darstellung von Nucleotid- und Aminosäuresequenzen in europäischen Patentanmeldungen*

Hat die in einer europäischen Patentanmeldung offenbarte Erfindung Nucleotid- oder Aminosäuresequenzen zum Gegenstand, so hat die Beschreibung ein Sequenzprotokoll zu enthalten, das dem im **Anhang** zu diesem Beschluß wiedergegebenen WIPO-Standard ST. 25 entspricht.

### Artikel 2

*Einreichung von Sequenzprotokollen auf elektronischem Datenträger*

(1) Zusätzlich ist ein Datenträger mit dem Sequenzprotokoll einzureichen, der dem im **Anhang** zu diesem Beschluß wiedergegebenen Standard entspricht.

(2) Wird ein Sequenzprotokoll nach seiner Einreichung beim Europäischen Patentamt berichtigt, so ist ein neuer Datenträger mit dem berichtigten Sequenzprotokoll einzureichen.

(3) Jedem Datenträger ist eine Erklärung des Anmelders beizufügen, daß die auf dem Datenträger gespeicherte Information mit dem schriftlichen Sequenzprotokoll übereinstimmt.

### Artikel 3

*Beseitigung von Mängeln*

(1) Wird den in Regel 27a EPÜ und in den Artikeln 1 und 2 dieses Beschlusses genannten Erfordernissen nicht entsprochen, so teilt die Eingangsstelle dies dem Anmelder mit und fordert ihn auf, die festgestellten Mängel innerhalb einer Frist von zwei Monaten zu beseitigen (Regel 41 (1) EPÜ).

## Decision of the President of the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings

The President of the European Patent Office, having regard to Article 10 EPC, Rules 27a and 49 EPC and Rule 13<sup>ter</sup> PCT, has decided as follows:

### Article 1

*Standardised presentation of nucleotide and amino acid sequences in European patent applications*

If the subject-matter of the invention disclosed in a European patent application is nucleotide or amino acid sequences, the description shall contain a sequence listing complying with WIPO Standard ST. 25 reproduced in the **annex** to this decision.

### Article 2

*Filing of sequence listings on electronic data carriers*

(1) In addition, a data carrier shall be filed containing the sequence listing and complying with the standard reproduced in the **annex** to this decision.

(2) If a sequence listing is corrected after being filed with the European Patent Office, a new data carrier containing the corrected sequence listing shall be filed.

(3) Each data carrier shall be accompanied by a statement by the applicant that the information recorded on the data carrier is identical to the written sequence listing.

### Article 3

*Rectification of deficiencies*

(1) If the requirements laid down in Rule 27a EPC and Articles 1 and 2 of this decision are not met, the Receiving Section shall inform the applicant accordingly and invite him to remedy the disclosed deficiencies within a period of two months (Rule 41(1) EPC).

## Décision du Président de l'Office européen des brevets, en date du 2 octobre 1998, relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences

Le Président de l'Office européen des brevets, vu l'article 10 CBE, les règles 27bis et 49 CBE ainsi que la règle 13<sup>ter</sup> PCT, décide :

### Article premier

*Présentation normalisée du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet européen*

Si l'invention exposée dans une demande de brevet européen a pour objet des séquences de nucléotides ou d'acides aminés, la description doit contenir un listage des séquences établi conformément à la norme ST. 25 de l'OMPI, reproduite dans l'**annexe** à la présente décision.

### Article 2

*Dépôt de listages des séquences sur support électronique de données*

(1) En outre, il y a lieu de déposer un support de données contenant un listage des séquences établi conformément à la norme reproduite dans l'**annexe** à la présente décision.

(2) Si un listage des séquences est rectifié après avoir été déposé auprès de l'Office européen des brevets, il y a lieu de déposer un nouveau support de données contenant le listage des séquences rectifié.

(3) Tout support de données doit être accompagné d'une déclaration du demandeur selon laquelle l'information figurant sur ce support est identique à celle que contient le listage écrit.

### Article 3

*Correction d'irrégularités*

(1) S'il n'est pas satisfait aux exigences visées à la règle 27bis CBE et aux articles premier et 2 de la présente décision, la section de dépôt le signale au demandeur et l'invite à remédier aux irrégularités constatées dans un délai de deux mois (règle 41(1) CBE).

(2) Die Frist wird einmal um höchstens zwei Monate verlängert, sofern der Anmelder dies vor Ablauf der Frist beantragt.

(3) Ist ein Datenträger beschädigt oder aus einem anderen Grund nicht lesbar, so teilt die Eingangsstelle dies dem Anmelder mit und fordert ihn auf, einen neuen, einwandfrei lesbaren Datenträger innerhalb einer von ihr zu bestimmenden Frist einzureichen.

#### Artikel 4

*Einreichung von Sequenzprotokollen beim EPA als Internationaler Recherchenbehörde*

In bezug auf internationale Anmeldungen nach dem PCT, für die nach Regel 5.2 PCT ein Sequenzprotokoll vorgeschrieben ist, verfährt das Europäische Patentamt als Internationale Recherchenbehörde nach Regel 13<sup>er</sup>.1 PCT und Abschnitten 208 und 513 mit Anhang C der Verwaltungsvorschriften zum PCT.

#### Artikel 5

*Einreichung von Sequenzprotokollen beim EPA als Bestimmungsamt oder als ausgewähltes Amt*

Bei Aufforderung zur Einreichung eines Sequenzprotokolls nach Regel 104b (3a) EPÜ ist Artikel 3 entsprechend anzuwenden.

#### Artikel 6

*Veröffentlichung von Sequenzprotokollen durch das EPA*

Ein am Anmeldetag der europäischen Patentanmeldung eingereichtes Sequenzprotokoll wird als Bestandteil der Beschreibung zusammen mit den Anmeldungsunterlagen veröffentlicht. Ein nach dem Anmeldetag eingereichtes oder berichtiges Sequenzprotokoll wird, wenn es vor dem Zeitpunkt vorliegt, an dem die technischen Vorbereitungen für die Veröffentlichung als abgeschlossen gelten, als Anlage zu den Anmeldungsunterlagen, andernfalls, bei Patenterteilung, als Anlage zur europäischen Patentschrift veröffentlicht.

(2) This period shall be extendable once only for a maximum of two months, provided the applicant requests such an extension prior to the expiry of the period.

(3) If a data carrier is damaged, or unreadable for some other reason, the Receiving Section shall inform the applicant accordingly and invite him to submit a flawless replacement within a period which it shall lay down.

#### Article 4

*Filing of sequence listings with the EPO as International Searching Authority*

With regard to international applications under the PCT, for which a sequence listing is prescribed under Rule 5.2 PCT, the European Patent Office shall act as International Searching Authority in accordance with Rule 13<sup>er</sup>.1 PCT and Sections 208 and 513 together with Annex C of the PCT Administrative Instructions.

#### Article 5

*Filing of sequence listings with the EPO as designated Office or elected Office*

Where an applicant is invited to file a sequence listing in accordance with Rule 104b(3a) EPC, Article 3 shall apply *mutatis mutandis*.

#### Article 6

*Publication of sequence listings by the EPO*

A sequence listing filed on the date of filing of the European patent application shall be published together with the application documents as part of the description. A sequence listing filed or corrected after the date of filing shall be published as an annex to the application documents, provided it is received before the date on which the technical preparations for publication are deemed to have been completed; otherwise, it shall be published as an annex to the European patent specification when the patent is granted.

(2) Le délai est prolongé une seule fois de deux mois au maximum, si le demandeur présente une requête en ce sens avant l'expiration du délai.

(3) Si un support de données est endommagé ou rendu illisible pour quelque autre raison, la section de dépôt le signale au demandeur et l'invite à déposer un nouveau support de données parfaitement lisible dans le délai qu'elle lui impartit.

#### Article 4

*Dépôt de listages des séquences auprès de l'OEB agissant en qualité d'administration chargée de la recherche internationale*

En ce qui concerne les demandes internationales déposées au titre du PCT pour lesquelles la règle 5.2 PCT prescrit un listage des séquences, l'Office européen des brevets, agissant en qualité d'administration chargée de la recherche internationale, procède conformément aux dispositions de la règle 13<sup>er</sup>.1 PCT et aux instructions administratives 208 et 513 du PCT, accompagnées de l'annexe C des Instructions administratives du PCT.

#### Article 5

*Dépôt de listages des séquences auprès de l'OEB agissant en tant qu'office désigné ou élu*

Lors de l'envoi d'une invitation à déposer un listage des séquences conformément à la règle 104ter (3bis) CBE, l'article 3 est applicable par analogie.

#### Article 6

*Publication par l'OEB de listages des séquences*

Un listage des séquences déposé à la date de dépôt de la demande de brevet européen est publié comme élément de la description avec les documents de la demande. Si un listage des séquences déposé ou rectifié après la date de dépôt est disponible avant la date à laquelle les préparatifs techniques entrepris en vue de la publication de la demande sont réputés achevés, il est publié en annexe aux documents de la demande, ou, si ce n'est pas le cas, lors de la délivrance du brevet, en annexe au fascicule du brevet européen.

**Artikel 7**

*Inkrafttreten*

Dieser Beschluß tritt am **1. Januar 1999** in Kraft und ersetzt den Beschluß vom 11. Dezember 1992, Beilage Nr. 2 zu ABI. EPA 12/1992, 1.

Geschehen zu München am  
2. Oktober 1998

*INGO KOBER*  
Präsident

**Article 7**

*Entry into force*

This decision shall enter into force on **1 January 1999** and shall replace the decision dated 11 December 1992, Supplement No. 2 to OJ EPO 12/1992, 1.

Done at Munich, 2 October 1998

*INGO KOBER*  
President

**Article 7**

*Entrée en vigueur*

La présente décision entre en vigueur le **1<sup>er</sup> janvier 1999** et remplace la décision du 11 décembre 1992, publiée dans le supplément n° 2 au Journal officiel 12/1992, 1.

Fait à Munich, le 2 octobre 1998

*INGO KOBER*  
Président

**ANHANG zum Beschluß des Präsidenten des Europäischen Patentamts vom 2. Oktober 1998 über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen**

**WIPO-Standard ST. 25**

**Standard für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in Patentanmeldungen**

Vom (ausführenden) PCIPI-Koordinierungsausschuß in seiner zweiundzwanzigsten Sitzung am 28. Mai 1998 angenommener Standard

Den Ämtern wird empfohlen, die Vorschriften des "Standards für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen nach dem Vertrag über die internationale Zusammenarbeit auf dem Gebiet des Patentwesens (PCT)", die in Anhang C zu den Verwaltungsvorschriften zum PCT enthalten sind, auch auf solche Patentanmeldungen anzuwenden, die keine internationalen Anmeldungen nach dem PCT sind, wobei zu beachten ist, daß bestimmte Vorschriften, die auf PCT-spezifische Verfahren und Erfordernisse zugeschnitten sind, auf Patentanmeldungen, die keine internationalen Anmeldungen nach dem PCT sind, möglicherweise keine Anwendung finden.\* Der Wortlaut des PCT-Standards ist nachfolgend wiedergegeben.

**ANNEX to the Decision of the President of the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings**

**WIPO Standard ST. 25**

**Standard for the presentation of nucleotide and amino acid sequence listings in patent applications**

Standard adopted by the PCIPI Executive Coordination Committee at its twenty-second session on May 28, 1998

It is recommended that Offices apply the provisions of the "Standard for the Presentation of Nucleotide and Amino Acid Sequence Listings in International Patent Applications Under the Patent Cooperation Treaty (PCT)" as set out in Annex C to the Administrative Instructions under the PCT, mutatis mutandis, to all patent applications other than the PCT international applications, noting that certain provisions specific to PCT procedures and requirements may not be applicable to patent applications other than PCT international applications.\* The text of that PCT Standard is reproduced on the following pages.

**ANNEXE à la décision du Président de l'Office européen des brevets, en date du 2 octobre 1998, relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences**

**Norme ST. 25 de l'OMPI**

**Norme relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet**

Norme adoptée par le Comité exécutif de coordination du PCIPI à sa vingt-deuxième session

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.\* Le texte de cette norme du PCT est reproduit ci-après.

\* Ist am 1. Juli 1998 die für ein Amt geltende nationale Gesetzgebung und Rechtspraxis nicht mit den Vorschriften der ersten beiden Sätze in Absatz 3 des "Standards für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen nach dem Vertrag über die internationale Zusammenarbeit auf dem Gebiet des Patentwesens (PCT)" vereinbar, so steht es diesem Amt frei, die betreffenden Vorschriften nicht zu befolgen, solange die Unvereinbarkeit besteht.

\* If, on July 1, 1998, the national law and practice applicable by an Office is not compatible with the provisions of the first two sentences of paragraph 3 of the "Standard for the Presentation of Nucleotide and Amino Acid Sequence Listings in International Patent Applications Under the Patent Cooperation Treaty (PCT)", that Office may choose not to follow those provisions for as long as that incompatibility continues.

\* Si, le 1<sup>er</sup> juillet 1998, la législation et la pratique nationales appliquées par un office ne sont pas compatibles avec les dispositions des deux premières phrases du paragraphe 3 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)", l'office peut choisir de ne pas appliquer ces dispositions tant que l'incompatibilité demeure.

## Anhang C zu den PCT-Verwaltungsvorschriften

Standard für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen nach dem PCT

### EINFÜHRUNG

1. Mit dem vorliegenden Standard soll auf eine standardisierte Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen hingewirkt werden. Anhand dieses Standards kann der Anmelder ein einheitliches Sequenzprotokoll erstellen, das von allen Anmeldeämtern, Internationalen Recherchenbehörden und mit der internationalen vorläufigen Prüfung beauftragten Behörden für die Zwecke der internationalen Phase und von allen Bestimmungsämtern und ausgewählten Ämtern für die Zwecke der nationalen Phase akzeptiert wird. Der Standard soll ferner eine genauere und qualitativ bessere Darstellung von Nucleotid- und Aminosäuresequenzen in internationalen Anmeldungen sicherstellen, die Darstellung und Verbreitung der Sequenzen zum Nutzen der Anmelder, der Öffentlichkeit und der Prüfer erleichtern, die Recherche von Sequenzdaten vereinfachen und den Austausch dieser Daten in elektronischer Form sowie ihre Aufnahme in computergestützte Datenbanken ermöglichen.

### DEFINITIONEN

2. Im Rahmen dieses Standards gelten folgende Definitionen:

i) Unter einem "Sequenzprotokoll" ist ein Teil der Beschreibung der Anmeldung in der eingereichten Fassung oder ein zur Anmeldung nachgereichtes Schriftstück zu verstehen, das die Nucleotid- und/oder Aminosäuresequenzen im einzelnen offenbart und sonstige verfügbare Angaben enthält.

ii) In das Protokoll dürfen nur unverzweigte Sequenzen von mindestens 4 Aminosäuren bzw. unverzweigte Sequenzen von mindestens 10 Nucleotiden aufgenommen werden. Verzweigte Sequenzen, Sequenzen mit weniger als 4 genau definierten Nucleotiden oder Aminosäuren sowie Sequenzen mit Nucleotiden oder Aminosäuren, die nicht in Anhang 2, Tabellen 1, 2, 3 und 4 aufgeführt sind, sind ausdrücklich ausgenommen.

iii) Unter "Nucleotiden" sind nur Nucleotide zu verstehen, die mittels der in Anhang 2, Tabelle 1 aufge-

## Annex C to the PCT Administrative Instructions

Standard for the presentation of nucleotide and amino acid sequence listings in international patent applications under the PCT

### INTRODUCTION

1. This Standard has been elaborated so as to provide standardization of the presentation of nucleotide and amino acid sequence listings in international patent applications. The Standard is intended to allow the applicant to draw up a single sequence listing which is acceptable to all receiving Offices, International Searching and Preliminary Examining Authorities for the purposes of the international phase, and to all designated and elected Offices for the purposes of the national phase. It is intended to enhance the accuracy and quality of presentations of nucleotide and amino acid sequences given in international applications, to make for easier presentation and dissemination of sequences for the benefit of applicants, the public and examiners, to facilitate searching of sequence data and to allow the exchange of sequence data in electronic form and the introduction of sequence data onto computerized databases.

### DEFINITIONS

2. For the purposes of this Standard:

(i) the expression "sequence listing" means a part of the description of the application as filed or a document filed subsequently to the application, which gives a detailed disclosure of the nucleotide and/or amino acid sequences and other available information;

(ii) sequences which are included are any unbranched sequences of four or more amino acids or unbranched sequences of ten or more nucleotides. Branched sequences, sequences with fewer than four specifically defined nucleotides or amino acids as well as sequences comprising nucleotides or amino acids other than those listed in Appendix 2, Tables 1, 2, 3 and 4, are specifically excluded from this definition;

(iii) "nucleotides" embrace only those nucleotides that can be represented using the symbols set forth

## Annexe C des Instructions administratives du PCT

Norme relative à la présentation du listage 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 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 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

fürten Symbole wiedergegeben werden können. Modifikationen wie z. B. methylierte Basen können nach der Anleitung in Anhang 2, Tabelle 2 beschrieben werden, sind aber in der Nucleotidsequenz nicht explizit auszuweisen.

iv) Unter "Aminosäuren" sind die in Anhang 2, Tabelle 3 aufgeführten gängigen L-Aminosäuren aus den natürlich vorkommenden Proteinen zu verstehen. Aminosäuresequenzen, die mindestens eine D-Aminosäure enthalten, fallen nicht unter diese Definition. Aminosäuresequenzen, die posttranslational modifizierte Aminosäuren enthalten, können mittels der in Anhang 2, Tabelle 3 aufgeführten Symbole wie die ursprünglich translatierte Aminosäure beschrieben werden, wobei die modifizierten Positionen wie z. B. Hydroxylierungen oder Glykosilierungen nach der Anleitung in Anhang 2, Tabelle 4 beschrieben werden, diese Modifikationen aber in der Aminosäuresequenz nicht explizit auszuweisen sind. Unter die vorstehende Definition fallen auch Peptide oder Proteine, die anhand der in Anhang 2, Tabelle 3 aufgeführten Symbole sowie einer an anderer Stelle aufgenommenen Beschreibung, die beispielsweise Aufschluß über ungewöhnliche Bindungen, Quervernetzungen (z. B. Disulfidbrücken) und "end caps", Nichtpeptidbindungen usw. gibt, als Sequenz wiedergegeben werden können.

v) Unter "Sequenzkennzahl" ist eine Zahl zu verstehen, die jeder im Protokoll aufgeführten Sequenz als SEQ ID NO zugewiesen wird.

vi) Die "numerische Kennzahl" ist eine dreistellige Zahl, die für ein bestimmtes Datenelement steht.

vii) Unter "sprachneutralem Vokabular" ist ein festes Vokabular zu verstehen, das im Sequenzprotokoll zur Wiedergabe der vom Hersteller einer Sequenzdatenbank vorgeschriebenen wissenschaftlichen Begriffe verwendet wird (dazu gehören wissenschaftliche Namen, nähere Bestimmungen und ihre Entsprechungen im festen Vokabular, die Symbole in Anhang 2, Tabellen 1, 2, 3 und 4 und die Merkmalschlüssel in Anhang 2, Tabellen 5 und 6).

viii) Unter "zuständiger Behörde" ist die Internationale Recherchenbehörde oder die mit der internationalen vorläufigen Prüfung beauftragte Behörde, die die internationale Recherche bzw. die internationale

in Appendix 2, Table 1. Modifications, for example, methylated bases, may be described as set forth in Appendix 2, Table 2, but shall not be shown explicitly in the nucleotide sequence;

(iv) "amino acids" are those L-amino acids commonly found in naturally occurring proteins and are listed in Appendix 2, Table 3. Those amino acid sequences containing at least one D-amino acid are not intended to be embraced by this definition. Any amino acid sequence that contains post-translationally modified amino acids may be described as the amino acid sequence that is initially translated using the symbols shown in Appendix 2, Table 3, with the modified positions, for example, hydroxylations or glycosylations, being described as set forth in Appendix 2, Table 4, but these modifications shall not be shown explicitly in the amino acid sequence. Any peptide or protein that can be expressed as a sequence using the symbols in Appendix 2, Table 3, in conjunction with a description elsewhere to describe, for example, abnormal linkages, cross-links (for example, disulfide bridge) and end caps, non-peptidyl bonds, etc., is embraced by this definition;

(v) "sequence identifier" is a unique integer that corresponds to the SEQ ID NO assigned to each sequence in the listing;

(vi) "numeric identifier" is a three-digit number which represents a specific data element;

(vii) "language-neutral vocabulary" is a controlled vocabulary used in the sequence listing that represents scientific terms as prescribed by sequence database providers (including scientific names, qualifiers and their controlled-vocabulary values, the symbols appearing in Appendix 2, Tables 1, 2, 3 and 4, and the feature keys appearing in Appendix 2, Tables 5 and 6);

(viii) "competent Authority" is the International Searching Authority that is to carry out the international search on the international application, or the International Preliminary Examining Authority that is to carry

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 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 de la recherche internationale ou l'administration chargée de l'examen préliminaire international pour la demande internationale

vorläufige Prüfung zu der internationalen Anmeldung durchführt, oder das Bestimmungsamt bzw. ausgewählte Amt zu verstehen, das mit der Bearbeitung der internationalen Anmeldung begonnen hat.

#### SEQUENZPROTOKOLL

3. Das Sequenzprotokoll im Sinne der Nummer 2 i) ist ans Ende der Anmeldung zu setzen, wenn es mit ihr zusammen eingereicht wird. Dieser Teil ist mit "Sequenzprotokoll" zu überschreiben, muß mit einer neuen Seite beginnen und sollte gesondert numeriert werden. Das Sequenzprotokoll ist Bestandteil der Beschreibung; vorbehaltlich Nummer 36 erübrigt es sich deshalb, die Sequenzen in der Beschreibung an anderer Stelle nochmals zu beschreiben.

4. Wird das Sequenzprotokoll im Sinne der Nummer 2 i) nicht zusammen mit der Anmeldung eingereicht, sondern als gesondertes Schriftstück nachgereicht (s. Nr. 37), so ist es mit der Überschrift "Sequenzprotokoll" und einer gesonderten Seitennummerierung zu versehen. Die in der Anmeldung in der eingereichten Fassung gewählte Numerierung der Sequenzen (s. Nr. 5) ist auch im nachgereichten Sequenzprotokoll beizubehalten.

5. Jeder Sequenz wird eine eigene Sequenzkennzahl zugeteilt. Die Kennzahlen beginnen mit 1 und setzen sich in aufsteigender Reihenfolge als ganze Zahlen fort. Gibt es zu einer Kennzahl keine Sequenz, so ist am Anfang der auf die SEQ ID NO folgenden Zeile unter der numerischen Kennzahl <400> der Code 000 anzugeben. Unter der numerischen Kennzahl <160> ist die Gesamtzahl der SEQ ID NOs anzugeben, und zwar unabhängig davon, ob im Anschluß an eine SEQ ID NO eine Sequenzkennzahl oder der Code 000 folgt.

6. In der Beschreibung, in den Ansprüchen und in den Zeichnungen der Anmeldung ist auf die im Sequenzprotokoll dargestellten Sequenzen mit "SEQ ID NO", gefolgt von der betreffenden Kennzahl, zu verweisen.

7. Für die Darstellung von Nucleotid- und Aminosäuresequenzen ist zumindest eine der folgenden drei Möglichkeiten zu wählen:

- i) nur Nucleotidsequenz
- ii) nur Aminosäuresequenz
- iii) Nucleotidsequenz zusammen mit der entsprechenden Aminosäuresequenz

out the international preliminary examination on the international application, or the designated/elected Office before which the processing of the international application has started.

#### SEQUENCE LISTING

3. The sequence listing as defined in paragraph 2(i) shall, where it is filed together with the application, be placed at the end of the application. This part shall be entitled "Sequence Listing," begin on a new page and preferably have independent page numbering. The sequence listing forms an integral part of the description; it is therefore unnecessary, subject to paragraph 36, to describe the sequences elsewhere in the description.

4. Where the sequence listing as defined in paragraph 2(i) is not contained in the application as filed but is a separate document furnished subsequently to the filing of the application (see paragraph 37), it shall be entitled "Sequence Listing" and shall have independent page numbering. The original numbering of the sequences (see paragraph 5) in the application as filed shall be maintained in the subsequently furnished sequence listing.

5. Each sequence shall be assigned a separate sequence identifier. The sequence identifiers shall begin with 1 and increase sequentially by integers. If no sequence is present for a sequence identifier, the code 000 should appear under numeric identifier <400>, beginning on the next line following the SEQ ID NO. The response for numeric identifier <160> shall include the total number of SEQ ID NOs, whether followed by a sequence or by the code 000.

6. In the description, claims or drawings of the application, the sequences represented in the sequence listing shall be referred to by the sequence identifier and preceded by "SEQ ID NO:".

7. Nucleotide and amino acid sequences should be represented by at least one of the following three possibilities:

- (i) a pure nucleotide sequence;
- (ii) a pure amino acid sequence;
- (iii) a nucleotide sequence together with its corresponding amino acid sequence.

en question, ou encore l'office désigné ou élu au sein duquel le traitement de la demande internationale a commencé.

#### LISTAGE DES SÉQUENCES

3. Le listage des séquences défini au paragraphe 2.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 "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 36, de décrire les séquences ailleurs dans la description.

4. Lorsque le listage des séquences défini au paragraphe 2.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 37), il doit s'intituler "Listage des séquences" et faire l'objet d'une pagination distincte. La numérotation des séquences (voir le paragraphe 5) figurant dans la demande telle qu'elle a été déposée doit être maintenue dans un listage des séquences remis ultérieurement.

5. A 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.

Bei einer Offenbarung im unter Ziffer iii genannten Format muß die Aminosäuresequenz als solche im Sequenzprotokoll mit einer eigenen Sequenzkennzahl gesondert offenbart werden.

For those sequences disclosed in the format specified in option (iii), above, the amino acid sequence must be disclosed separately in the sequence listing as a pure amino acid sequence with a separate integer sequence identifier.

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.

## NUCLEOTIDSEQUENZEN

### *Zu verwendende Symbole*

8. Nucleotidsequenzen sind nur anhand eines Einzelstrangs in Richtung vom 5'-Ende zum 3'-Ende von links nach rechts wiederzugeben. Die Begriffe 3' und 5' werden in der Sequenz nicht dargestellt.

9. Die Basen einer Nucleotidsequenz sind anhand der einbuchstabigen Codes für Nucleotidsequenzzeichen darzustellen. Es dürfen nur die in Anhang 2, Tabelle 1 aufgeführten Kleinbuchstaben verwendet werden.

10. Modifizierte Basen sind in der Sequenz selbst wie die entsprechenden unmodifizierten Basen oder mit "n" wiederzugeben, wenn die modifizierte Base zu den in Anhang 2, Tabelle 2 aufgeführten gehört; die Modifikation ist im Merkmalsteil des Sequenzprotokolls anhand der Codes in Anhang 2, Tabelle 2 näher zu beschreiben. Diese Codes dürfen in der Beschreibung oder im Merkmalsteil des Sequenzprotokolls, nicht jedoch in der Sequenz selbst verwendet werden (s. auch Nr. 32). Das Symbol "n" steht immer nur für ein einziges unbekanntes oder modifiziertes Nucleotid.

### *Zu verwendendes Format*

11. Bei Nucleotidsequenzen sind höchstens 60 Basen pro Zeile – mit einem Leerraum zwischen jeder Gruppe von 10 Basen – aufzuführen.

12. Die Basen einer Nucleotidsequenz (einschließlich Intronen) sind jeweils in Zehnergruppen aufzuführen; dies gilt nicht für die codierenden Teile der Sequenz. Bleiben am Ende nicht-codierender Teile einer Sequenz weniger als 10 Basen übrig, so sind sie zu einer Gruppe zusammenzufassen und durch einen Leerraum von angrenzenden Gruppen zu trennen.

13. Die Basen der codierenden Teile einer Nucleotidsequenz sind als Triplets (Codonen) aufzuführen.

## NUCLEOTIDE SEQUENCES

### *Symbols to Be Used*

8. A nucleotide sequence shall be presented only by a single strand, in the 5'-end to 3'-end direction from left to right. The terms 3' and 5' shall not be represented in the sequence.

9. The bases of a nucleotide sequence shall be represented using the one-letter code for nucleotide sequence characters. Only lower case letters in conformity with the list given in Appendix 2, Table 1, shall be used.

10. Modified bases shall be represented as the corresponding unmodified bases or as "n" in the sequence itself if the modified base is one of those listed in Appendix 2, Table 2, and the modification shall be further described in the feature section of the sequence listing, using the codes given in Appendix 2, Table 2. These codes may be used in the description or the feature section of the sequence listing but not in the sequence itself (see also paragraph 32). The symbol "n" is the equivalent of only one unknown or modified nucleotide.

### *Format to Be Used*

11. A nucleotide sequence shall be listed with a maximum of 60 bases per line, with a space between each group of 10 bases.

12. The bases of a nucleotide sequence (including introns) shall be listed in groups of 10 bases, except in the coding parts of the sequence. Leftover bases, fewer than 10 in number at the end of non-coding parts of a sequence, should be grouped together and separated from adjacent groups by a space.

13. The bases of the coding parts of a nucleotide sequence shall be listed as triplets (codons).

## 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. Die Zählung der Nucleotidbasen beginnt bei der ersten Base der Sequenz mit 1. Von hier aus ist die gesamte Sequenz in 5'-3'-Richtung fortlaufend durchzuzählen. Am rechten Rand ist jeweils neben der Zeile mit den einbuchstabigen Codes für die Basen die Nummer der letzten Base dieser Zeile anzugeben. Die vorstehend beschriebene Zählweise für Nucleotidsequenzen gilt auch für Nucleotidsequenzen mit ringförmiger Konfiguration, wobei allerdings die Bestimmung des ersten Nucleotids der Sequenz dem Anmelder überlassen bleibt.

15. Eine Nucleotidsequenz, die aus einem oder mehreren nichtbenachbarten Abschnitten einer größeren Sequenz oder aus Abschnitten verschiedener Sequenzen besteht, ist als gesonderte Sequenz mit eigener Sequenzkennzahl zu numerieren. Sequenzen mit einer oder mehreren Lücken sind als mehrere gesonderte Sequenzen mit eigenen Sequenzkennzahlen zu numerieren, wobei die Zahl der gesonderten Sequenzen der Zahl der jeweils zusammenhängenden Sequenzdatenreihen entspricht.

#### AMINOSÄURESEQUENZEN

##### *Zu verwendende Symbole*

16. Die Aminosäuren einer Protein- oder Peptidsequenz sind in Richtung von der Amino- zur Carboxylgruppe von links nach rechts aufzuführen, wobei die Amino- und Carboxylgruppen in der Sequenz nicht darzustellen sind.

17. Die Aminosäuren sind anhand des dreibuchstabigen Codes mit großem Anfangsbuchstaben entsprechend der Liste in Anhang 2, Tabelle 3 darzustellen. Eine Aminosäuresequenz, die einen Leerraum oder interne Terminatorsymbole (z. B. "Ter", "\*" oder ".") enthält, ist nicht als eine einzige Aminosäuresequenz, sondern als getrennte Aminosäuresequenzen darzustellen (s. Nr. 22).

18. Modifizierte und seltene Aminosäuren sind in der Sequenz selbst wie die entsprechenden unmodifizierten Aminosäuren oder mit "Xaa" wiederzugeben, wenn sie zu den in Anhang 2, Tabelle 4 aufgeführten gehören und die Modifikation im Merkmalsteil des Sequenzprotokolls anhand der Codes in Anhang 2, Tabelle 4 näher beschrieben wird. Diese Codes dürfen in der Beschreibung oder im Merkmalsteil des

14. The enumeration of the nucleotide shall start at the first base of the sequence with number 1. It shall be continuous through the whole sequence in the direction 5' to 3'. It shall be marked in the right margin, next to the line containing the one-letter codes for the bases, and giving the number of the last base of that line. The enumeration method for nucleotide sequences set forth above remains applicable to nucleotide sequences that are circular in configuration, with the exception that the designation of the first nucleotide of the sequence may be made at the option of the applicant.

15. A nucleotide sequence that is made up of one or more non-contiguous segments of a larger sequence or of segments from different sequences shall be numbered as a separate sequence, with a separate sequence identifier. A sequence with a gap or gaps shall be numbered as a plurality of separate sequences with separate sequence identifiers, with the number of separate sequences being equal in number to the number of continuous strings of sequence data.

#### AMINO ACID SEQUENCES

##### *Symbols to Be Used*

16. The amino acids in a protein or peptide sequence shall be listed in the amino to carboxy direction from left to right. The amino and carboxy groups shall not be represented in the sequence.

17. The amino acids shall be represented using the three-letter code with the first letter as a capital and shall conform to the list given in Appendix 2, Table 3. An amino acid sequence that contains a blank or internal terminator symbols (for example, "Ter" or "\*" or ".") may not be represented as a single amino acid sequence, but shall be presented as separate amino acid sequences (see paragraph 22).

18. Modified and unusual amino acids shall be represented as the corresponding unmodified amino acids or as "Xaa" in the sequence itself if the modified amino acid is one of those listed in Appendix 2, Table 4, and the modification shall be further described in the feature section of the sequence listing, using the codes given in Appendix 2, Table 4. These codes may be used in the description or the feature section of

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

Sequenzprotokolls, nicht jedoch in der Sequenz selbst verwendet werden (s. auch Nr. 32). Das Symbol "Xaa" steht immer nur für eine einzige unbekannte oder modifizierte Aminosäure.

the sequence listing but not in the sequence itself (see also paragraph 32). The symbol "Xaa" is the equivalent of only one unknown or modified amino acid.

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.

#### *Zu verwendendes Format*

19. Bei Protein- oder Peptidsequenzen sind höchstens 16 Aminosäuren pro Zeile mit einem Leerraum zwischen den einzelnen Aminosäuren aufzuführen.

#### *Format to Be Used*

19. A protein or peptide sequence shall be listed with a maximum of 16 amino acids per line, with a space provided between each amino acid.

#### *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. Die den Codonen der codierenden Teile einer Nucleotidsequenz entsprechenden Aminosäuren sind unmittelbar unter den jeweiligen Codonen anzugeben. Wird ein Codon durch ein Intron aufgespalten, so ist das Aminosäuresymbol unter dem Teil des Codons anzugeben, der zwei Nucleotide enthält.

20. Amino acids corresponding to the codons in the coding parts of a nucleotide sequence shall be placed immediately under the corresponding codons. Where a codon is split by an intron, the amino acid symbol should be given below the portion of the codon containing two nucleotides.

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. Die Zählung der Aminosäuren beginnt bei der ersten Aminosäure der Sequenz mit 1. Fakultativ können die dem reifen Protein vorausgehenden Aminosäuren wie beispielsweise Präsequenzen, Prosequenzen, Präprosequenzen und Signalsequenzen, soweit vorhanden, mit negativem Vorzeichen numeriert werden, wobei die Rückwärtszählung mit der Aminosäure vor Nummer 1 beginnt. Null (0) wird nicht verwendet, wenn Aminosäuren zur Abgrenzung gegen das reife Protein mit negativem Vorzeichen numeriert werden. Die Nummern sind im Abstand von jeweils 5 Aminosäuren unter der Sequenz anzugeben. Die vorstehend beschriebene Zählweise für Aminosäuresequenzen gilt auch für Aminosäuresequenzen mit ringförmiger Konfiguration, wobei allerdings die Bestimmung der ersten Aminosäure der Sequenz dem Anmelder überlassen bleibt.

21. The enumeration of amino acids shall start at the first amino acid of the sequence, with number 1. Optionally, the amino acids preceding the mature protein, for example pre-sequences, pro-sequences, pre-pro-sequences and signal sequences, when present, may have negative numbers, counting backwards starting with the amino acid next to number 1. Zero (0) is not used when the numbering of amino acids uses negative numbers to distinguish the mature protein. It shall be marked under the sequence every five amino acids. The enumeration method for amino acid sequences set forth above remains applicable for amino acid sequences that are circular in configuration, with the exception that the designation of the first amino acid of the sequence may be made at the option of the applicant.

21. L'énumération des acides aminés doit commencer par le premier acide aminé de la séquence, qui portera le numéro 1. A titre facultatif, les acides aminés précédant la protéine mature, par exemple les pré-sequences, les prosequences et les pré-prosequences 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. Eine Aminosäuresequenz, die aus einem oder mehreren nichtbenachbarten Abschnitten einer größeren Sequenz oder aus Abschnitten verschiedener Sequenzen besteht, ist als gesonderte Sequenz mit eigener Sequenzkennzahl zu numerieren. Sequenzen mit einer oder mehreren Lücken sind als mehrere gesonderte Sequenzen mit eigenen Sequenzkennzahlen zu numerieren, wobei die Zahl der gesonderten Sequenzen der Zahl der jeweils zusammenhängenden Sequenzdatenreihen entspricht.

22. An amino acid sequence that is made up of one or more non-contiguous segments of a larger sequence or of segments from different sequences shall be numbered as a separate sequence, with a separate sequence identifier. A sequence with a gap or gaps shall be numbered as a plurality of separate sequences with separate sequence identifiers, with the number of separate sequences being equal in number to the number of continuous strings of sequence data.

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.

SONSTIGE VERFÜGBARE ANGABEN  
IM SEQUENZPROTOKOLL

23. Die Angaben sind im Sequenzprotokoll in der Reihenfolge anzugeben, in der sie in der Liste der numerischen Kennzahlen der Datenelemente in Anhang 1 aufgeführt sind.

24. Für die Angaben im Sequenzprotokoll sind nur die in Anhang 1 aufgeführten numerischen Kennzahlen, nicht aber die dazugehörigen Beschreibungen zu verwenden. Die Angaben müssen unmittelbar auf die numerische Kennzahl folgen; im Sequenzprotokoll brauchen nur diejenigen numerischen Kennzahlen angegeben zu werden, zu denen auch Angaben vorliegen. Die einzigen beiden Ausnahmen zu dieser Vorschrift bilden die numerischen Kennzahlen <220> und <300>, die für die Rubrik "Merkmal" bzw. "Veröffentlichungsangaben" stehen und mit den Angaben unter den numerischen Kennzahlen <221> bis <223> bzw. <301> bis <313> zusammenhängen. Werden unter diesen numerischen Kennzahlen im Sequenzprotokoll Angaben zu den Merkmalen oder zur Veröffentlichung gemacht, so sollte auch die numerische Kennzahl <220> bzw. <300> aufgeführt, die dazugehörige Rubrik aber nicht ausgefüllt werden. Generell sollte zwischen den numerischen Kennzahlen eine Leerzeile eingefügt werden, wenn sich die an erster oder zweiter Position der numerischen Kennzahl stehende Ziffer ändert. Eine Ausnahme von dieser allgemeinen Regel bildet die numerische Kennzahl <310>, der keine Leerzeile vorausgehen darf. Auch vor jeder Wiederholung der numerischen Kennzahl ist eine Leerzeile einzufügen.

*Obligatorische Datenelemente*

25. In das Sequenzprotokoll sind außerdem vor der eigentlichen Nucleotid- und/oder Aminosäuresequenz die folgenden in Anhang 1 definierten Angaben (obligatorische Datenelemente) aufzunehmen:

<110> Name des Anmelders  
<120> Bezeichnung der Erfindung  
<160> Anzahl der SEQ ID NOs  
<210> SEQ ID NO: x  
<211> Länge  
<212> Art  
<213> Organismus  
<400> Sequenz

OTHER AVAILABLE INFORMATION IN  
THE SEQUENCE LISTING

23. The order of the items of information in the sequence listings shall follow the order in which those items are listed in the list of numeric identifiers of data elements as defined in Appendix 1.

24. Only numeric identifiers of data elements as defined in Appendix 1 shall be used for the presentation of the items of information in the sequence listing. The corresponding numeric identifier descriptions shall not be used. The provided information shall follow immediately after the numeric identifier while only those numeric identifiers for which information is given need appear on the sequence listing. Two exceptions to this requirement are numeric identifiers <220> and <300>, which serve as headers for "Feature" and "Publication Information," respectively, and are associated with information in numeric identifiers <221> to <223> and <301> to <313>, respectively. When feature and publication information is provided in the sequence listing under those numeric identifiers, numeric identifiers <220> and <300>, respectively, should be included, but left blank. Generally, a blank line shall be inserted between numeric identifiers when the digit in the first or second position of the numeric identifier changes. An exception to this general rule is that no blank line should appear preceding numeric identifier <310>. Additionally, a blank line shall precede any repeated numeric identifier.

*Mandatory Data Elements*

25. The sequence listing shall include, in addition to and immediately preceding the actual nucleotide and/or amino acid sequence, the following items of information defined in Appendix 1 (mandatory data elements):

<110> Applicant name  
<120> Title of invention  
<160> Number of SEQ ID NOs  
<210> SEQ ID NO: x  
<211> Length  
<212> Type  
<213> Organism  
<400> Sequence

AUTRES RENSEIGNEMENTS  
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.

*Eléments 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

Ist der Name des Anmelders (numerische Kennzahl <110>) nicht in Buchstaben des lateinischen Alphabets geschrieben, so ist er – im Wege der Transliteration oder der Übersetzung ins Englische – auch in lateinischen Buchstaben anzugeben.

Where the name of the applicant (numeric identifier <110>) is written in characters other than those of the Latin alphabet, it shall also be indicated in characters of the Latin alphabet either as a mere transliteration or through translation into English.

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.

Die Datenelemente mit Ausnahme der Angaben unter den numerischen Kennzahlen <110>, <120> und <160> sind für jede im Sequenzprotokoll aufgeführte Sequenz zu wiederholen. Gibt es zu einer Sequenzkennzahl keine Sequenz, so müssen nur die Datenelemente unter den numerischen Kennzahlen <210> und <400> angegeben werden (s. Nr. 5 und SEQ ID NO: 4 in dem in Anhang 3 dieses Standards enthaltenen Beispiel).

The data elements, except those under numeric identifiers <110>, <120> and <160>, shall be repeated for each sequence included in the sequence listing. Only the data elements under numeric identifiers <210> and <400> are mandatory if no sequence is present for a sequence identifier (see paragraph 5, above, and SEQ ID NO: 4 in the example depicted in Appendix 3 of this Standard).

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. In Sequenzprotokolle, die zusammen mit der dazugehörigen Anmeldung oder vor Vergabe einer Anmeldeungsnummer eingereicht werden, ist neben den unter Nummer 25 genannten Datenelementen auch das folgende aufzunehmen:

26. In addition to the data elements identified in paragraph 25, above, when a sequence listing is filed at the same time as the application to which it pertains or at any time prior to the assignment of an application number, the following data element shall be included in the sequence listing:

26. Outre les éléments de données indiqués au paragraphe 25, 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> Aktenzeichen

<130> File reference

<130> Référence du dossier

27. In Sequenzprotokolle, die auf Anforderung einer zuständigen Behörde oder nach Vergabe einer Anmeldeungsnummer eingereicht werden, sind neben den unter Nummer 25 genannten Datenelementen auch die folgenden aufzunehmen:

27. In addition to the data elements identified in paragraph 25, above, when a sequence listing is filed in response to a request from a competent Authority or at any time following the assignment of an application number, the following data elements shall be included in the sequence listing:

27. Outre les éléments de données énumérés au paragraphe 25, 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> Vorliegende Patentanmeldung  
<141> Anmeldetag der vorliegenden Anmeldung

<140> Current patent application  
<141> Current filing date

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

28. In Sequenzprotokolle, die zu einer Anmeldung eingereicht werden, die die Priorität einer früheren Anmeldung in Anspruch nimmt, sind neben den unter Nummer 25 genannten Datenelementen auch die folgenden aufzunehmen:

28. In addition to the data elements identified in paragraph 25, above, when a sequence listing is filed relating to an application which claims the priority of an earlier application, the following data elements shall be included in the sequence listing:

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> Frühere Patentanmeldung

<150> Earlier patent application

<150> Demande de brevet antérieure

<151> Anmeldetag der früheren Anmeldung

<151> Earlier application filing date

<151> Date de dépôt de la demande antérieure

29. Wird in der Sequenz "n", "Xaa" oder eine modifizierte Base oder modifizierte/seltene L-Aminosäure aufgeführt, so müssen die folgenden Datenelemente angegeben werden:

<220> Merkmal  
<221> Name/Schlüssel  
<222> Lage  
<223> Sonstige Informationen

30. Ist der Organismus (numerische Kennzahl <213>) eine "künstliche Sequenz" oder "unbekannt", so müssen die folgenden Datenelemente angegeben werden:

<220> Merkmal  
<223> Sonstige Angaben

#### *Fakultative Datenelemente*

31. Alle in Anhang 1 definierten Datenelemente, die unter den Nummern 25 bis 30 nicht erwähnt sind, sind fakultativ (fakultative Datenelemente).

#### *Angabe von Merkmalen*

32. Merkmale, die (unter der numerischen Kennzahl <220>) zu einer Sequenz angegeben werden, sind durch die in Anhang 2, Tabellen 5 und 6<sup>1</sup> aufgeführten "Merkmal-schlüssel" zu beschreiben.

#### *Freier Text*

33. Unter "freiem Text" ist eine verbale Beschreibung der Eigenschaften der Sequenz ohne Verwendung des sprachneutralen Vokabulars im Sinne der Nummer 2 (vii) unter der numerischen Kennzahl <223> (Sonstige Angaben) zu verstehen.

34. Der freie Text sollte sich auf einige kurze, für das Verständnis der Sequenz unbedingt notwendige Begriffe beschränken. Er sollte in englischer Sprache für jedes Datenelement nicht länger als 4 Zeilen sein und höchstens 65 Buchstaben pro Zeile umfassen. Alle weiteren Angaben sind in den Hauptteil der Beschreibung in der dort verwendeten Sprache aufzunehmen.

35. Freier Text sollte in englischer Sprache abgefaßt sein.

36. Enthält das Sequenzprotokoll, das Bestandteil der Beschreibung ist, freien Text, so ist dieser im Hauptteil der Beschreibung in der dort verwen-

29. If "n" or "Xaa" or a modified base or modified/unusual L-amino acid is used in the sequence, the following data elements are mandatory:

<220> Feature  
<221> Name/key  
<222> Location  
<223> Other information

30. If the organism (numeric identifier <213>) is "Artificial Sequence" or "Unknown," the following data elements are mandatory:

<220> Feature  
<223> Other information

#### *Optional Data Elements*

31. All data elements defined in Appendix 1, not mentioned in paragraphs 25 to 30, above, are optional (optional data elements).

#### *Presentation of Features*

32. When features of sequences are presented (that is, numeric identifier <220>), they shall be described by the "feature keys" set out in Appendix 2, Tables 5 and 6.<sup>1</sup>

#### *Free Text*

33. "Free text" is a wording describing characteristics of the sequence under numeric identifier <223> (Other information) which does not use language-neutral vocabulary as referred to in paragraph 2(vii).

34. The use of free text shall be limited to a few short terms indispensable for the understanding of the sequence. It shall not exceed four lines with a maximum of 65 characters per line for each given data element, when written in English. Any further information shall be included in the main part of the description in the language thereof.

35. Any free text should preferably be in the English language.

36. Where the sequence listing part of the description contains free text, any such free text shall be repeated in the main part of the description in

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

#### *Eléments 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.

#### *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.

36. Lorsque, dans la description, la partie réservée au listage des séquences contient du texte libre, celui-ci doit être répété dans la partie

<sup>1</sup> Diese Tabellen enthalten Auszüge aus den Merkmals-tabellen "DDBJ/EMBL/GenBank Feature Table" (Nucleotidsequenzen) und "SWISS PROT Feature Table" (Aminosäuresequenzen).

<sup>1</sup> These tables contain extracts of the DDBJ/EMBL/GenBank Feature Table (nucleotide sequences) and the SWISS PROT Feature Table (amino acid sequences).

<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).

deten Sprache zu wiederholen. Es wird empfohlen, den in der Sprache des Hauptteils der Beschreibung abgefaßten freien Text in einen besonderen Abschnitt der Beschreibung mit der Überschrift "Sequenzprotokoll – freier Text" aufzunehmen.

#### NACHGERECHTES SEQUENZPROTOKOLL

37. Alle Sequenzprotokolle, die nicht zusammen mit der Anmeldung, sondern später eingereicht werden, dürfen nicht über den Offenbarungsgehalt der Anmeldung in der eingereichten Fassung hinausgehen; ihnen ist eine entsprechende Erklärung beizufügen. Die zur Anmeldung nachgereichten Sequenzprotokolle dürfen also nur die Sequenzen enthalten, die in der Anmeldung in der eingereichten Fassung offenbart worden sind.

38. Sequenzprotokolle, die nicht zusammen mit der Anmeldung eingereicht worden sind, sind nicht Bestandteil der Anmeldung. Da die Regeln 13<sup>ter</sup>, 26.3 und 91 sowie der Artikel 34 PCT Anwendung finden, besteht jedoch – vorbehaltlich der einschlägigen Bestimmungen – die Möglichkeit, Sequenzprotokolle, die zusammen mit der Anmeldung eingereicht wurden, im Wege der Mängelbeseitigung gemäß Regel 13<sup>ter</sup> oder 26.3 PCT und – bei Vorliegen offensichtlicher Fehler – gemäß Regel 91 PCT zu berichtigen oder gemäß Artikel 34 PCT zu ändern sowie ferner im Wege der Änderung der Anmeldung gemäß Artikel 34 PCT ein Sequenzprotokoll nachzureichen.

#### SEQUENZPROTOKOLL IN MASCHINENLESBARER FORM

39. Das in der Anmeldung enthaltene Sequenzprotokoll ist zusätzlich auch in maschinenlesbarer Form einzureichen, wenn die zuständige Behörde dies verlangt.

40. Das zusätzlich eingereichte maschinenlesbare Sequenzprotokoll muß mit dem geschriebenen Protokoll identisch sein und ist zusammen mit einer Erklärung folgenden Wortlauts einzureichen: "Die in maschinenlesbarer Form aufgezeichneten Angaben sind mit dem geschriebenen Sequenzprotokoll identisch."

the language thereof. It is recommended that the free text in the language of the main part of the description be put in a specific section of the description called "Sequence Listing Free Text."

#### SUBSEQUENTLY FURNISHED SEQUENCE LISTING

37. Any sequence listing which is not contained in the application as filed but which is furnished subsequently shall not go beyond the disclosure in the application as filed and shall be accompanied by a statement to that effect. This means that a sequence listing furnished subsequently to the filing of the application shall contain only those sequences that were disclosed in the application as filed.

38. Any sequence listing not contained in the application as filed does not form part of the application. However, the provisions of PCT Rules 13<sup>ter</sup>, 26.3 and 91 and PCT Article 34 would apply, so that it may be possible, subject to the applicable provisions, for a sequence listing contained in the application as filed to be corrected under PCT Rules 13<sup>ter</sup> or 26.3, rectified under PCT Rule 91 (in the case of an obvious error), or amended under PCT Article 34, or for a sequence listing to be submitted under PCT Article 34 as an amendment to the application.

#### COMPUTER READABLE FORM OF THE SEQUENCE LISTING

39. A copy of the sequence listing shall also be submitted in computer readable form, in addition to the sequence listing as contained in the application, whenever this is required by the competent Authority.

40. Any sequence listing in computer readable form submitted in addition to the written sequence listing shall be identical to the written sequence listing and shall be accompanied by a statement that "the information recorded in computer readable form is identical to the written sequence listing."

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 "texte libre du listage des séquences".

#### LISTAGE DES SÉQUENCES DÉPOSÉ ULTÉRIEUREMENT

37. 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.

38. Tout listage des séquences ne figurant pas dans la demande telle qu'elle a été déposée ne fait pas partie de la demande. Toutefois, les dispositions des règles 13<sup>ter</sup>, 26.3 et 91 du règlement d'exécution du PCT et celles de l'article 34 du PCT s'appliquent normalement, de telle sorte qu'il est possible, sous réserve des dispositions applicables, qu'un listage des séquences figurant dans la demande telle qu'elle a été déposée fasse l'objet d'une correction en vertu de la règle 13<sup>ter</sup> ou 26.3 du règlement d'exécution du PCT, d'une rectification en vertu de la règle 91 du règlement d'exécution du PCT (dans le cas d'une erreur évidente) ou d'une modification en vertu de l'article 34 du PCT, ou qu'un listage des séquences soit soumis en vertu de l'article 34 du PCT à titre de modification de la demande.

#### LISTAGE DES SÉQUENCES SOUS UNE FORME DÉCHIFFRABLE PAR ORDINATEUR

39. Outre le listage des séquences figurant dans la demande, une copie de ce même listage doit être fournie sous une forme déchiffrable par ordinateur chaque fois que l'administration compétente l'exige.

40. 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."

41. Das gesamte ausdrückbare Exemplar des Sequenzprotokolls muß in einer einzigen Datei enthalten sein, die nach Möglichkeit auf einer einzigen Diskette oder einem sonstigen von der zuständigen Behörde akzeptierten elektronischen Datenträger aufgezeichnet sein soll. Diese Datei ist mittels der IBM<sup>2</sup>-Codetabellen (Code Page) 437, 932<sup>3</sup> oder einer kompatiblen Codetabelle zu codieren. Eine Codetabelle, wie sie z. B. für japanische, chinesische, kyrillische, arabische, griechische oder hebräische Schriftzeichen benötigt wird, gilt als kompatibel, wenn sie das lateinische Alphabet und die arabischen Ziffern denselben Hexadezimalpositionen zuordnet wie die genannten Codetabellen.

42. Die maschinenlesbare Fassung soll mit spezieller Software wie PatentIn oder anderen kundenspezifischen Computerprogrammen erstellt werden; sie kann mit beliebigen Mitteln erstellt werden, solange das Sequenzprotokoll auf einer Diskette oder einem anderen von der zuständigen Behörde akzeptierten elektronischen Träger eingereicht wird und mit einem Betriebssystem für Personal Computer gelesen werden kann, das von der zuständigen Behörde akzeptiert wird.

43. Bei Verwendung von Disketten ist eine Datenkompression erlaubt, sofern die komprimierte Datei in einem selbstextrahierenden Format erstellt worden ist, das sich auf einem von der zuständigen Behörde akzeptierten Betriebssystem für Personal Computer selbst dekomprimiert.

44. Auf der Diskette oder einem anderen von der zuständigen Behörde akzeptierten elektronischen Träger ist ein Etikett fest anzubringen, auf dem von Hand oder mit der Maschine in Blockschrift der Name des Anmelders, die Bezeichnung der Erfindung, ein Aktenzeichen, der Zeitpunkt der Aufzeichnung der Daten, das Computerbetriebssystem und der Name der zuständigen Behörde eingetragen sind.

45. Wird die Diskette oder ein anderes von der zuständigen Behörde akzeptiertes elektronisches Medium erst nach dem Tag der Anmeldung

41. The entire printable copy of the sequence listing shall be contained within one electronic file preferably on a single diskette or any other electronic medium that is acceptable to the competent Authority. The file recorded on the diskette or any other electronic medium that is acceptable to the competent Authority shall be encoded using IBM<sup>2</sup> Code Page 437, IBM Code Page 932 or a compatible code page. A compatible code page, as would be required for, for example, Japanese, Chinese, Cyrillic, Arabic, Greek or Hebrew characters, is one that assigns the Roman alphabet and numerals to the same hexadecimal positions as do the specified code pages.

42. The computer readable form shall preferably be created by dedicated software such as PatentIn or other custom computer programs; it may be created by any means, as long as the sequence listing on a submitted diskette or any other electronic medium that is acceptable to the competent Authority is readable under a Personal Computer Operating system that is acceptable to the competent Authority.

43. File compression is acceptable when using diskette media, so long as the compressed file is in a self-extracting format that will decompress on a Personal Computer Operating system that is acceptable to the competent Authority.

44. The diskette or any other electronic medium that is acceptable to the competent Authority shall have a label permanently affixed thereto on which has been hand-printed, in block capitals or typed, the name of the applicant, the title of the invention, a reference number, the date on which the data were recorded, the computer operating system and the name of the competent Authority.

45. If the diskette or any other electronic medium that is acceptable to the competent Authority is submitted after the date of filing of an applica-

41. 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'administration compétente. 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 ou une page de code compatible. 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.

42. Le listage déchiffirable par ordinateur doit, de préférence, être créé par un logiciel spécialisé tel que PatentIn ou d'autres programmes informatiques personnalisés; il peut être créé par tout autre moyen dans la mesure où le listage des séquences figurant sur la disquette ou sur tout autre support électronique admis par l'administration compétente est déchiffirable sous un système d'exploitation d'ordinateur individuel lui aussi admis par l'administration compétente.

43. La compression d'un fichier est admise lorsque le support est une disquette, dans la mesure où le fichier compressé se présente sous un format auto-extractible qui se décompressera sur un système d'exploitation d'ordinateur personnel admis par l'administration compétente.

44. Sur la disquette ou sur tout autre support électronique admis par l'administration compétente doit être apposée 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, le système d'exploitation informatique et le nom de l'administration compétente.

45. Si la disquette ou tout autre support électronique admis par l'administration compétente est fournie après la date de dépôt de la

<sup>2</sup> IBM ist eine für die International Business Machine Corporation, Vereinigte Staaten von Amerika, eingetragene Handelsmarke.

<sup>3</sup> Die genannten Codetabellen gelten für Personal Computer praktisch als Standard.

<sup>2</sup> IBM is a registered trademark of International Business Machine Corporation, United States of America.

<sup>3</sup> The specified code pages are *de facto* standards for personal computers.

<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.

eingereicht, so sind auf den Etiketten auch der Anmeldetag und die Anmelde­nummer anzugeben.

46. Jeder Berichtigung des geschriebenen Sequenzprotokolls nach Regel 13<sup>ter</sup>.1 (a) i) oder 26.3 PCT, jeder Berichtigung eines offensichtlichen Fehlers im geschriebenen Sequenzprotokoll nach Regel 91 PCT und jeder nach Artikel 34 PCT vorgenommenen Änderung, die sich auf ein geschriebenes Sequenzprotokoll bezieht, ist ein maschinenlesbares Sequenzprotokoll beizufügen, in dem diese Berichtigung oder Änderung berücksichtigt worden ist.

tion, the labels shall also include the filing date of the application and the application number.

46. Any correction of the written sequence listing which is submitted under PCT Rules 13<sup>ter</sup>.1(a)(i) or 26.3, any rectification of an obvious error in the written sequence listing which is submitted under PCT Rule 91, or any amendment which includes a written sequence listing and which is submitted under PCT Article 34, shall be accompanied by a computer readable form of the sequence listing including any such correction, rectification or amendment.

demande, cette date et le numéro de la demande doivent aussi figurer sur les étiquettes.

46. Toute correction – en vertu des règles 13<sup>er</sup>.1.a)i) ou 26.3 du règlement d'exécution du PCT – du listage des séquences présenté par écrit, toute rectification – en vertu de la règle 91 du règlement d'exécution du PCT – d'une erreur évidente dans le listage des séquences présenté par écrit ou toute modification – en vertu de l'article 34 du PCT – qui comprend un listage des séquences présenté par écrit doit être accompagnée du listage des séquences sous une forme déchiffrable par ordinateur, comprenant la correction, la rectification ou la modification en question.



**Anhang 1****Numerische Kennzahlen**

In Sequenzprotokollen, die zu Anmeldungen eingereicht werden, dürfen nur die nachstehenden numerischen Kennzahlen verwendet werden. Die Überschriften der nachstehenden Datenelementrubriken dürfen in den Sequenzprotokollen nicht erscheinen.

Die numerischen Kennzahlen der obligatorischen Datenelemente, d. h. der Datenelemente, die in alle Sequenzprotokolle aufgenommen werden müssen (s. Nr. 25 dieses Standards: Kennziffern 110, 120, 160, 210, 211, 212, 213 und 400), und die numerischen Kennzahlen der Datenelemente, die in den in diesem Standard genannten Fällen aufgenommen werden müssen (s. Nr. 26,

27, 28, 29 und 30 dieses Standards: Kennzahlen 130, 140, 141, 150 und 151 sowie 220 bis 223), sind mit dem Buchstaben "O" gekennzeichnet.

Die numerischen Kennzahlen der fakultativen Datenelemente (s. Nr. 31 dieses Standards) sind mit dem Buchstaben "F" gekennzeichnet.

Numerische Kennzahl	Numerische Kennzahl Beschreibung	Obligatorisch (O) oder fakultativ (F)	Bemerkungen
<110>	Name des Anmelders	O	Ist der Name des Anmelders nicht in lateinischen Buchstaben geschrieben, so muß er – im Wege der Transliteration oder der Übersetzung ins Englische – auch in lateinischen Buchstaben angegeben werden.
<120>	Bezeichnung der Erfindung	O	
<130>	Aktenzeichen	O in den Fällen nach Nr. 26 Standard	Siehe Nr. 26 Standard.
<140>	Vorliegende Patentanmeldung	O in den Fällen nach Nr. 27 Standard	Siehe Nr. 27 Standard; die vorliegende Patentanmeldung ist zu kennzeichnen durch den zweibuchstabigen Code nach dem WIPO-Standard ST. 3, gefolgt von der Anmeldungsnummer (in dem Format, das von der Behörde für gewerblichen Rechtsschutz verwendet wird, bei der diese Patentanmeldung eingereicht wird) oder – bei internationalen Anmeldungen – von der internationalen Anmeldungsnummer.
<141>	Anmeldetag der vorliegenden Anmeldung	O in den Fällen nach Nr. 27 Standard	Siehe Nr. 27 Standard; das Datum ist entsprechend dem WIPO-Standard ST. 2 anzugeben (CCYY MM DD).
<150>	Frühere Patentanmeldung	O in den Fällen nach Nr. 28 Standard	Siehe Nr. 28 Standard; die frühere Patentanmeldung ist zu kennzeichnen durch den zweibuchstabigen Code entsprechend dem WIPO-Standard ST. 3, gefolgt von der Anmeldungsnummer (in dem Format, das von der Behörde für gewerblichen Rechtsschutz verwendet wird, bei der die frühere Patentanmeldung eingereicht wurde) oder – bei internationalen Anmeldungen – von der internationalen Anmeldungsnummer.
<151>	Anmeldetag der früheren Anmeldung	O in den Fällen nach Nr. 28 Standard	Siehe Nr. 28 Standard; das Datum ist entsprechend dem WIPO-Standard ST. 2 anzugeben (CCYY MM DD).
<160>	Anzahl der SEQ ID NOs	O	
<170>	Software	F	
<210>	Angaben zu SEQ ID NO: x	O	Anzugeben ist eine ganze Zahl, die die SEQ ID NO darstellt.
<211>	Länge	O	Sequenzlänge, ausgedrückt als Anzahl der Basenpaare oder Aminosäuren
<212>	Art	O	Art des in SEQ ID NO: x sequenzierten Moleküls, und zwar entweder DNA, RNA oder PRT; enthält eine Nucleotidsequenz

			sowohl DNA- als auch RNA-Fragmente, so ist "DNA" anzugeben; zusätzlich ist das kombinierte DNA-/RNA-Molekül im Merkmalsteil unter <220> bis <223> näher zu beschreiben.
<213>	Organismus	O	Gattung/Art (d. h. wissenschaftlicher Name), "künstliche Sequenz" oder "unbekannt"
<220>	Merkmal	O in den Fällen nach Nr. 29 und 30 Standard	Freilassen; siehe Nr. 29 und 30 Standard; Beschreibung biologisch signifikanter Stellen in der Sequenz gemäß SEQ ID NO: x) (kann je nach der Zahl der angegebenen Merkmale mehrmals vorkommen).
<221>	Name/Schlüssel	O in den Fällen nach Nr. 29 Standard	Siehe Nr. 29 Standard; es dürfen nur die in Anhang 2, Tabelle 5 oder 6 beschriebenen Schlüssel verwendet werden.
<222>	Lage	O in den Fällen nach Nr. 29 Standard	Siehe Nr. 29 Standard; <ul style="list-style-type: none"> <li>- von (Nummer der ersten Base/Aminosäure des Merkmals)</li> <li>- bis (Nummer der letzten Base/Aminosäure des Merkmals)</li> <li>- Basenpaare (Ziffern verweisen auf die Positionen der Basenpaare in einer Nucleotidsequenz)</li> <li>- Aminosäuren (Ziffern verweisen auf die Positionen der Aminosäurereste in einer Aminosäuresequenz)</li> <li>- Angabe, ob sich das Merkmal auf dem zum Strang des Sequenzprotokolls komplementären Strang befindet</li> </ul>
<223>	Sonstige Angaben	O in den Fällen nach Nr. 29 und 30 Standard	Siehe Nr. 29 und 30 Standard; sonstige relevante Angaben, wobei sprachneutrales Vokabular oder freier Text (möglichst in englischer Sprache) zu verwenden ist; freier Text ist im Hauptteil der Beschreibung in der dort verwendeten Sprache zu wiederholen (s. Nr. 36 Standard); enthält die Sequenz eine der in Anhang 2, Tabellen 2 und 4 aufgeführten modifizierten Basen oder modifizierten/seltenen L-Aminosäuren, so ist für diese Base oder Aminosäure das zugehörige Symbol aus Anhang 2, Tabellen 2 und 4 zu verwenden.
<300>	Veröffentlichungsangaben	F	Freilassen; dieser Abschnitt ist für jede relevante Veröffentlichung zu wiederholen.
<301>	Verfasser	F	
<302>	Titel	F	Titel der Veröffentlichung
<303>	Zeitschrift	F	Name der Zeitschrift, in der die Daten veröffentlicht wurden
<304>	Band	F	Band der Zeitschrift, in dem die Daten veröffentlicht wurden
<305>	Heft	F	Nummer des Hefts der Zeitschrift, in dem die Daten veröffentlicht wurden
<306>	Seiten	F	Seiten der Zeitschrift, auf denen die Daten veröffentlicht wurden

<307>	Datum	F	Datum der Zeitschrift, an dem die Daten veröffentlicht wurden; Angabe nach Möglichkeit entsprechend dem WIPO-Standard ST. 2 (CCYY MM DD)
<308>	Eingangsnummer in der Datenbank	F	von der Datenbank zugeteilte Eingangsnummer einschließlich Datenbankbezeichnung
<309>	Datenbank-Eingabedatum	F	Datum der Eingabe in die Datenbank; Angabe entsprechend dem WIPO-Standard ST. 2 (CCYY MM DD)
<310>	Dokumentennummer	F	Nummer des Dokuments, nur bei Patentdokumenten; die vollständige Nummer hat nacheinander folgendes zu enthalten: den zweibuchstabigen Code entsprechend dem WIPO-Standard ST. 3, die Veröffentlichungsnummer entsprechend dem WIPO-Standard ST. 6 und den Code für die Dokumentenart nach dem WIPO-Standard ST. 16
<311>	Anmeldetag	F	Anmeldetag des Dokuments, nur bei Patentdokumenten; Angabe entsprechend WIPO-Standard ST. 2 (CCYY MM DD)
<312>	Veröffentlichungsdatum	F	Datum der Veröffentlichung des Dokuments; nur bei Patentdokumenten; Angabe entsprechend WIPO-Standard ST. 2 (CCYY MM DD)
<313>	Relevante Reste in SEQ ID NO: x von bis	F	
<400>	Sequenz	O	SEQ ID NO: x sollte in der der Sequenz vorausgehenden Zeile hinter der numerischen Kennzahl stehen (s. Anlage 3).

## Anhang 2

### Symbole für Nucleotide und Aminosäuren und Merkmalstabellen

Tabelle 1: Liste der Nucleotide

Symbol	Bedeutung	Ableitung der Bezeichnung
a	a	<u>A</u> denin
g	g	<u>G</u> uanin
c	c	<u>C</u> ytosin
t	t	<u>T</u> hymine
u	u	<u>U</u> racil
r	g oder a	Pu <u>r</u> in
y	t/u oder c	Pyri <u>m</u> idin
m	a oder c	A <u>m</u> ino
k	g oder t/u	<u>K</u> eto
s	g oder c	<u>s</u> tarke Bindungen 3 H-Brücken
w	a oder t/u	schwache (e: <u>w</u> weak) Bindungen 2 H-Brücken
b	b oder c oder t/u	nicht a
d	a oder g oder t/u	nicht c
h	a oder c oder t/u	nicht g
v	a oder g oder c	nicht t, nicht u
n	a oder g oder c oder t/u, unbekannt oder sonstige	beliebig (e: <u>a</u> ny)

Tabelle 2: Liste der modifizierten Nucleotide

Symbol	Bedeutung
ac4c	4-Acetylcytidin
chm5u	5-(Carboxyhydroxymethyl)uridin
cm	2'-O-Methylcytidin
cmnm5s2u	5-Carboxymethylaminomethyl-2-thiouridin
cmnm5u	5-Carboxymethylaminomethyluridin
d	Dihydrouridin
fm	2'-O-Methylpseudouridin
gal q	beta, D-Galactosylqueuosin
gm	2'-O-Methylguanosin
i	Inosin
i6a	N6-Isopentenyladenosin
m1a	1-Methyladenosin
m1f	1-Methylpseudouridin
m1g	1-Methylguanosin
m1i	1-Methylinosin
m22g	2,2-Dimethylguanosin
m2a	2-Methyladenosin
m2g	2-Methylguanosin
m3c	3-Methylcytidin
m5c	5-Methylcytidin
m6a	N6-Methyladenosin
m7g	7-Methylguanosin
mam5u	5-Methylaminomethyluridin
mam5s2u	5-Methoxyaminomethyl-2-thiouridin
man q	beta, D-Mannosequeuosin
mcm5s2u	5-Methoxycarbonylmethyl-2-thiouridin
mcm5u	5-Methoxycarbonylmethyluridin
mo5u	5-Methoxyuridin
ms2i6a	2-Methylthio-N6-isopentenyladenosin
ms2t6a	N-((9-beta-D-Ribofuranosyl-2-methylthiopurin-6-yl) carbamoyl) threonin
mt6a	N-((9-beta-D-Ribofuranosylpurin-6-yl)N-methylcarbamoyl)threonin
mv	Uridin-5-oxyessigsäuremethylester
o5u	Uridin-5-oxyessigsäure (v)
osyw	Wybutoxosin
p	Pseudouridin

q	Qeuosin
s2c	2-Thiocytidin
s2t	5-Methyl-2-thiouridin
s2u	2-Thiouridin
s4u	4-Thiouridin
t	5-Methyluridin
t6a	N-((9-beta-D-Ribofuranosylpurin-6-yl) carbamoyl)threonin
tm	2'-O-Methyl-5-methyluridin
um	2'-O-Methyluridin
yw	Wybutosin
x	3-(3-Amino-3-carboxypropyl)uridin, (acp3)u

Tabelle 3: Liste der Aminosäuren

Symbol	Bedeutung
Ala	Alanin
Cys	Cystein
Asp	Asparaginsäure
Glu	Glutaminsäure
Phe	Phenylalanin
Gly	Glycin
His	Histidin
Ile	Isoleucin
Lys	Lysin
Leu	Leucin
Met	Methionin
Asn	Asparagin
Pro	Prolin
Gln	Glutamin
Arg	Arginin
Ser	Serin
Thr	Threonin
Val	Valin
Trp	Tryptophan
Tyr	Tyrosin
Asx	Asp oder Asn
Glx	Glu oder Gln
Xaa	unbekannt oder sonstige

Tabelle 4: Liste der modifizierten und seltenen Aminosäuren

Symbol	Bedeutung
Aad	2-Aminoadipinsäure
bAad	3-Aminoadipinsäure
bAla	beta-Alanin, beta-Aminopropionsäure
Abu	2-Aminobuttersäure
4Abu	4-Aminobuttersäure, Piperidinsäure
Acp	6-Aminocapronsäure
Ahe	2-Aminoheptansäure
Aib	2-Aminoisobuttersäure
bAib	3-Aminoisobuttersäure
Apm	2-Aminopimelinsäure
Dbu	2,4-Diaminobuttersäure
Des	Desmosin
Dpm	2,2'-Diaminopimelinsäure
Dpr	2,3-Diaminopropionsäure
EtGly	N-Ethylglycin
EtAsn	N-Ethylasparagin
Hyl	Hydroxylysin
aHyl	allo-Hydroxylysin
3Hyp	3-Hydroxyprolin
4Hyp	4-Hydroxyprolin
Ide	Isodesmosin
alle	allo-Isoleucin
MeGly	N-Methylglycin, Sarkosin
Melle	N-Methylisoleucin
MeLys	6-N-Methyllysin
MeVal	N-Methylvalin
Nva	Norvalin
Nle	Norleucin
Orn	Ornithin

Tabelle 5: Liste der Merkmalschlüssel zu Nucleotidsequenzen

Schlüssel	Beschreibung
allele	Ein verwandtes Individuum oder ein verwandter Stamm enthält stabile alternative Formen desselben Gens und unterscheidet sich an dieser (und vielleicht an anderer) Stelle von der vorliegenden Sequenz.
attenuator	1. Region einer DNA, in der die Beendigung der Transkription reguliert wird und die Expression einiger bakterieller Operons gesteuert wird 2. zwischen dem Promotor und dem ersten Strukturgen liegender Sequenzabschnitt, der eine partielle Beendigung der Transkription bewirkt
C_region	konstante Region der leichten und schweren Immunglobulinketten und der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; enthält je nach Kette ein oder mehrere Exons
CAAT_signal	CAAT-Box; Teil einer konservierten Sequenz, der etwa 75 Basenpaare stromaufwärts vom Startpunkt der eukaryontischen Transkriptionseinheiten liegt und an der RNA-Polymerase-Bindung beteiligt sein kann; Konsensussequenz = GG (C oder T) CAATCT
CDS	codierende Sequenz; Sequenz von Nucleotiden, die mit der Sequenz der Aminosäuren in einem Protein übereinstimmt (beinhaltet Stopcodon); Merkmal schließt eine mögliche Translation der Aminosäure ein
conflict	Unabhängige Determinierungen "derselben" Sequenz unterscheiden sich an dieser Stelle oder in dieser Region voneinander.
D-loop	D-Schleife; Region innerhalb der mitochondrialen DNA, in der sich ein kürzeres RNA-Stück mit einem Strang der doppelsträngigen DNA paart und dabei den ursprünglichen Schwesterstrang in dieser Region verdrängt; dient auch zur Beschreibung der Verdrängung einer einzelsträngigen Region des DNA-Doppelstrangs durch einen einzelsträngigen Eindringling bei der durch ein recA-Protein ausgelösten Reaktion
D-segment	Diversity-Region der schweren Kette von Immunglobulin und der Beta-Kette eines T-Zell-Rezeptors
enhancer	eine als Cis-Element wirkende Sequenz, die die Aktivität (einiger) eukaryontischer Promotoren verstärkt und in beliebiger Richtung und Position zum Promotor (stromaufwärts oder -abwärts) funktioniert
exon	Region des Genoms, die für einen Teil der gespleißten mRNA codiert; kann 5'UTR, alle CDSs und 3'UTR enthalten
GC_signal	GC-Box; eine konservierte, GC-reiche Region stromaufwärts vom Startpunkt der eukaryontischen Transkriptionseinheiten, die in mehreren Kopien und in beiden Richtungen vorkommen kann; Konsensussequenz = GGGCGG
gene	biologisch signifikante Region, die als Gen bezeichnet wird und einen Namen trägt
iDNA	intervenierende DNA; DNA, die durch verschiedene Arten der Rekombination eliminiert wird
intron	DNA-Abschnitt, der transkribiert, aber beim Zusammenspleißen der ihn umrahmenden Sequenzen (Exons) aus dem Transkript wieder herausgeschnitten wird
J_segment	J-Kette (Verbindungskette) zwischen den leichten und den schweren Immunglobulinketten und den Alpha-, Beta- und Gamma-Ketten der T-Zell-Rezeptoren
LTR	lange Terminalwiederholung, eine an den beiden Enden einer gegebenen Sequenz direkt wiederholte Sequenz, wie sie für Retroviren typisch ist
mat_peptide	für ein reifes Peptid oder Protein codierende Sequenz; Sequenz, die im Anschluß an eine posttranslationale Modifizierung für das reife oder endgültige Peptid- oder Proteinprodukt codiert; schließt im Gegensatz zur entsprechenden CDS das Stopcodon nicht ein
misc_binding	Stelle in einer Nucleinsäure, die einen anderen Teil, der nicht durch einen anderen Bindungsschlüssel (primer_bind oder protein_bind) beschrieben werden kann, kovalent oder nicht kovalent bindet
misc_difference	Die Merkmalsequenz unterscheidet sich von der im Eintrag und kann nicht durch einen anderen Unterscheidungsschlüssel (conflict, unsure, old_sequence, mutation, variation, allele bzw. modified_base) beschrieben werden.



misc_feature	biologisch signifikante Region, die nicht durch einen anderen Merkmalschlüssel beschrieben werden kann; neues oder seltenes Merkmal
misc_recomb	Stelle, an der ein allgemeiner, ortsspezifischer oder replikativer Rekombinationsvorgang stattfindet, bei dem die DNA-Doppelhelix aufgebrochen und wieder zusammengefügt wird, und die nicht durch andere Rekombinationsschlüssel (iDNA oder Virion) oder den betreffenden Herkunftsschlüssel (/insertions_seq, /transposon, /proviral) beschrieben werden kann
misc_RNA	Transkript oder RNA-Produkt, das nicht durch andere RNA-Schlüssel (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 oder snRNA) beschrieben werden kann
misc_signal	Region, die ein Signal enthält, das die Genfunktion oder -expression steuert oder ändert, und die nicht durch andere Signalschlüssel (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator oder rep_origin) beschrieben werden kann
misc_structure	Sekundär-, Tertiär- oder sonstige Struktur oder Konformation, die nicht durch andere Strukturschlüssel (stem_loop oder D-loop) beschrieben werden kann
modified_base	Das angegebene Nucleotid ist ein modifiziertes Nucleotid und ist durch das angegebene Molekül (ausgedrückt durch die modifizierte Base) zu ersetzen.
mRNA	messenger-RNA (Boten-RNA); enthält eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon) und eine 3'-nichttranslatierte Region (3'UTR)
mutation	Ein verwandter Stamm weist an dieser Stelle eine plötzliche, erbliche Sequenzveränderung auf.
N_region	zusätzliche Nucleotide, die zwischen neu geordnete Immunglobulinabschnitte eingefügt werden
old_sequence	Die vorliegende Sequenz stellt eine geänderte Version der früher an dieser Stelle befindlichen Sequenz dar.
polyA_signal	Erkennungsregion, die zur Endonuclease-Spaltung eines RNA-Transkripts mit anschließender Polyadenylierung nötig ist; Konsensussequenz: AATAAA
polyA_site	Stelle auf einem RNA-Transkript, an der durch posttranslationale Polyadenylierung Adeninreste eingefügt werden
precursor_RNA	noch nicht gereifte RNA-Spezies; kann eine am 5'-Ende abzuschneidende Region (5'Clip), eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron), eine 3'-nichttranslatierte Region (3'UTR) und eine am 3'-Ende abzuschneidende Region (3'Clip) enthalten
prim_transcript	primäres (ursprüngliches, nicht prozessiertes) Transkript; enthält eine am 5'-Ende abzuschneidende Region (5'Clip), eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron), eine 3'-nichttranslatierte Region (3'UTR) und eine am 3'-Ende abzuschneidende Region (3'Clip)
primer_bind	nichtkovalente Primer-Bindungsstelle für die Initiierung der Replikation, Transkription oder reversen Transkription; enthält eine oder mehrere Stellen für synthetische Elemente, z. B. PCR-Primerelemente
promoter	Region auf einem DNA-Molekül, die an der Bindung der RNA-Polymerase beteiligt ist, die die Transkription initiiert
protein_bind	nichtkovalente Protein-Bindungsstelle auf der Nucleinsäure
RBS	ribosomale Bindungsstelle
repeat_region	Genomregion mit repetitiven Einheiten
repeat_unit	einzelnes Repeat (repetitive Einheit)
rep_origin	Replikationsursprung; Startpunkt der Duplikation der Nucleinsäure, durch die zwei identische Kopien entstehen
rRNA	reife ribosomale RNA; RNA-Komplex des Ribonucleoprotein-Partikels (Ribosom), der Aminosäuren zu Proteinen zusammenfügt
S_region	Switch-Region der schweren Immunglobulinketten; beteiligt am Umbau schwerer DNA-Ketten, der zur Expression einer anderen Immunglobulin-Klasse aus derselben B-Zelle führt

satellite	viele tandemartig hintereinandergeschaltete (identische oder verwandte) Repeats einer kurzen basischen repetitiven Einheit; viele davon unterscheiden sich in der Basenzusammensetzung oder einer anderen Eigenschaft vom Genomdurchschnitt und können so von der Hauptmasse der genomischen DNA abgetrennt werden
scRNA	kleine cytoplasmische RNA; eines von mehreren kleinen cytoplasmischen RNA-Molekülen im Cytoplasma und (manchmal) im Zellkern eines Eukaryonten
sig_peptide	für ein Signalpeptid codierende Sequenz; Sequenz, die für eine N-terminale Domäne eines sekretorischen Proteins codiert; diese Domäne spielt bei der Anheftung des naszierenden Polypeptids an die Membran eine Rolle; Leader-Sequenz
snRNA	kleine Kern-RNA; eine der vielen kleinen RNA-Formen, die nur im Zellkern vorkommen; einige der snRNAs spielen beim Spleißen oder bei anderen RNA-verarbeitenden Reaktionen eine Rolle
source	Bezeichnet die biologische Herkunft des genannten Sequenzabschnitts; die Angabe dieses Schlüssels ist obligatorisch; jeder Eintrag muß mindestens einen Herkunftsschlüssel aufweisen, der die gesamte Sequenz umfaßt; es dürfen zu jeder Sequenz auch mehrere Herkunftsschlüssel angegeben werden.
stem_loop	Haarnadelschleife; eine Doppelhelix-Region, die durch Basenpaarung zwischen benachbarten (invertierten) komplementären Sequenzen in einem RNA- oder DNA-Einzelstrang entsteht
STS	Sequence Tagged Site; kurze, nur als Einzelkopie vorkommende DNA-Sequenz, die einen Kartierungspunkt auf dem Genom bezeichnet und durch PCR ermittelt werden kann; eine Region auf dem Genom kann durch Bestimmung der Reihenfolge der STSs kartiert werden
TATA_signal	TATA-Box; Goldberg-Hogness-Box; ein konserviertes AT-reiches Septamer, das sich rund 25 Basenpaare vor dem Startpunkt jeder eukaryontischen RNA-Polymerase-II-Transkriptionseinheit befindet und bei der Positionierung des Enzyms für eine korrekte Initiation eine Rolle spielen kann; Konsensussequenz = TATA(A oder T)A(A oder T)
terminator	DNA-Sequenz, die entweder am Ende des Transkripts oder neben einer Promotor-Region liegt und bewirkt, daß die RNA-Polymerase die Transkription beendet; kann auch die Bindungsstelle eines Repressor-Proteins sein
transit_peptide	für Transitpeptid codierende Sequenz; Sequenz, die für eine N-terminale Domäne eines im Zellkern codierten Organellen-Proteins codiert; diese Domäne ist an der posttranslationalen Einschleusung des Proteins in die Organelle beteiligt
tRNA	reife transfer-RNA, ein kleines RNA-Molekül (75–85 Basen lang), das die Translation einer Nucleinsäure-Sequenz in eine Aminosäure-Sequenz vermittelt
unsure	Der Autor kennt die Sequenz in dieser Region nicht genau.
V_region	variable Region der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; codiert für das variable Aminoende; kann aus V-, D-, N- und J-Abschnitten bestehen
V-segment	variabler Abschnitt der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; codiert für den Großteil der variablen Region (V_region) und die letzten Aminosäuren des Leader-Peptids
variation	Ein verwandter Stamm enthält stabile Mutationen desselben Gens (z. B. RFLPs, Polymorphismen usw.), die sich an dieser (und möglicherweise auch an anderer) Stelle von der vorliegenden Sequenz unterscheiden.
3'clip	3'-äußerste Region eines Precursor-Transkripts, die bei der Prozessierung abgeschnitten wird
3'UTR	Region am 3'-Ende eines reifen Transkripts (nach dem Stopcodon), die nicht in ein Protein translatiert wird
5'clip	5'-äußerste Region eines Precursor-Transkripts, die bei der Prozessierung abgeschnitten wird
5'UTR	Region am 5'-Ende eines reifen Transkripts (vor dem Initiationscodon), die nicht in ein Protein translatiert wird
-10_signal	Pribnow-Box; konservierte Region rund 10 Basenpaare stromaufwärts vom Startpunkt der bakteriellen Transkriptionseinheiten, die bei der Bindung der RNA-Polymerase eine Rolle spielt; Konsensussequenz = TAtAAT
-35_signal	konserviertes Hexamer rund 35 Basenpaare stromaufwärts vom Startpunkt der bakteriellen Transkriptionseinheiten; Konsensussequenz = TTGACa [ ] oder TGTTGACA [ ]

Tabelle 6: Liste der Merkmalschlüssel zu Proteinsequenzen

Schlüssel	Beschreibung
CONFLICT	In den einzelnen Unterlagen ist von verschiedenen Sequenzen die Rede.
VARIANT	Den Angaben der Autoren zufolge gibt es Sequenzvarianten.
VARSPLIC	Beschreibung von Sequenzvarianten, die durch alternatives Spleißen entstanden sind
MUTAGEN	experimentell veränderte Stelle
MOD_RES	posttranslationale Modifikation eines Rests
ACETYLATION	N-terminale oder sonstige
AMIDATION	in der Regel am C-Terminus eines reifen aktiven Peptids
BLOCKED	unbestimmte Gruppe, die das N- oder C-terminale Ende blockiert
FORMYLATION	des N-terminalen Methionin
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	von Asparagin, Asparaginsäure, Prolin oder Lysin
METHYLATION	in der Regel von Lysin oder Arginin
PHOSORYLATION	von Serin, Threonin, Tyrosin, Asparaginsäure oder Histidin
PYRROLIDONE	N-terminales Glutamat, das ein internes cyclisches Lactam gebildet hat
CARBOXYLIC ACID SULFATATION	in der Regel von Tyrosin
LIPID	kovalente Bindung einer Lipidhälfte
MYRISTATE	Myristat-Gruppe, die durch eine Amidbindung an den N-terminalen Glycin-Rest der reifen Form eines Proteins oder an einen internen Lysin-Rest gebunden ist
PALMITATE	Palmitat-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest oder durch eine Esterbindung an einen Serin- oder Threonin-Rest gebunden ist
FARNESYL	Farnesyl-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest gebunden ist
GERANYL-GERANYL	Geranyl-geranyl-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest gebunden ist
GPI-ANCHOR	Glykosyl-phosphatidylinositol-(GPI-)Gruppe, die an die alpha-Carboxylgruppe des C-terminalen Rests der reifen Form eines Proteins gebunden ist
N-ACYL DIGLYCERIDE	N-terminales Cystein der reifen Form eines prokaryontischen Lipoproteins mit einer amidgebundenen Fettsäure und einer Glyceryl-Gruppe, an die durch Esterbindungen zwei Fettsäuren gebunden sind
DISULFID	Disulfidbindung; den "VON"- und den "BIS"-Endpunkt bilden die beiden Reste, die durch eine ketteninterne Disulfidbindung verbunden sind; sind der "VON"- und der "BIS"-Endpunkt identisch, liegt die Disulfidbindung innerhalb der Kette, und die Art der Quervernetzung ist im Beschreibungsfeld anzugeben
THIOLEST	Thiolesterbindung; den "VON"- und den "BIS"-Endpunkt bilden die beiden Reste, die durch die Thiolesterbindung verbunden sind
THIOETH	Thioetherbindung; den "VON"- und den "BIS"-Endpunkt bilden die beiden Reste, die durch die Thioetherbindung verbunden sind
CARBOHYD	Glykosylierungs-Stelle; die Art des Kohlenhydrats (sofern bekannt) ist im Beschreibungsfeld anzugeben
METAL	Bindungsstelle für ein Metallion; die Art des Metalls ist im Beschreibungsfeld anzugeben

BINDING	Bindungsstelle für eine beliebige chemische Gruppe (Coenzym, prosthetische Gruppe usw.); die Art der Gruppe ist im Beschreibungsfeld anzugeben
SIGNAL	Bereich einer Signalsequenz (Präpeptid)
TRANSIT	Bereich eines Transit-Peptids (mitochondriales, chloroplastisches oder für Microbodies)
PROPEP	Bereich eines Propeptids
CHAIN	Bereich einer Polypeptid-Kette im reifen Protein
PEPTIDE	Bereich eines freigesetzten aktiven Peptids
DOMAIN	Bereich einer wichtigen Domäne auf der Sequenz; die Art dieser Domäne ist im Beschreibungsfeld anzugeben
CA_BIND	Bereich einer Calcium-bindenden Region
DNA_BIND	Bereich einer DNA-bindenden Region
NP_BIND	Bereich einer Nucleotidphosphat-bindenden Region; die Art des Nucleotidphosphats ist im Beschreibungsfeld anzugeben
TRANSMEM	Bereich einer Transmembran-Region
ZN_FING	Bereich einer Zink-Finger-Region
SIMILAR	Grad der Ähnlichkeit mit einer anderen Proteinsequenz; im Beschreibungsfeld sind genaue Angaben über diese Sequenz zu machen
REPEAT	Bereich einer internen Sequenzwiederholung
HELIX	Sekundärstruktur: Helices, z. B. Alpha-Helix, 3(10)-Helix oder Pi-Helix
STRAND	Sekundärstruktur: Beta-Strang, z. B. durch Wasserstoff-Brückenbindungen stabilisierter Beta-Strang, oder Rest in einer isolierten Beta-Brücke
TURN	Sekundärstruktur: Schleife, z. B. durch Wasserstoff-Brückenbindungen stabilisierte Schleife (3-, 4- oder 5-Schleife)
ACT_SITE	Aminosäure(n), die bei der Aktivität eines Enzyms mitwirkt (mitwirken)
SITE	irgendeine andere wichtige Stelle auf der Sequenz
INIT_MET	die Sequenz beginnt bekanntermaßen mit einem Start-Methionin
NON_TER	Der Rest am Sequenzanfang oder -ende ist nicht der Terminalrest; steht er an der Position 1, so bedeutet das, daß diese nicht der N-Terminus des vollständigen Moleküls ist; steht er an letzter Position, so ist diese Position nicht der C-Terminus des vollständigen Moleküls; für diesen Schlüssel gibt es kein Beschreibungsfeld.
NON_CONS	nicht aufeinanderfolgende Reste; zeigt an, daß zwei Reste in einer Sequenz nicht aufeinanderfolgen, sondern daß zwischen ihnen einige nichtsequenzierte Reste liegen
UNSURE	Unsicherheiten in der Sequenz; mit diesem Schlüssel werden Regionen einer Sequenz beschrieben, bei der sich der Autor bezüglich der Sequenzzuweisung nicht sicher ist

### Anhang 3

#### Muster eines Sequenzprotokolls

```

<110>      Smith, John; Smithgene Inc.
<120>      Beispiel für ein Sequenzprotokoll
<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  cggccctct  cgcgctcctc  tcgcgctct  ctctcgctct  cctctcgctc           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
    
```

<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> Künstliche Sequenz

<220>  
 <223> Peptid, das nach Größe und Polarität als Verbindung zwischen den  
 Alpha- und Beta-Ketten des Proteins XYZ konzipiert ist

<400> 3

Met Val Asn Leu Glu Pro Met His Thr Glu Ile  
 1 5 10

<210> 4  
 <400> 4  
 000

**Appendix 1****Numeric identifiers**

Only numeric identifiers as defined below may be used in sequence listings submitted in applications. The text of the data element headings given below shall not be included in the sequence listings.

Numeric identifiers of mandatory data elements, that is, data elements which must be included in all sequence listings (see paragraph 25 of this Standard: items 110, 120, 160, 210, 211, 212, 213 and 400) and numeric identifiers of data elements which must be included in circumstances specified in this Standard (see paragraphs 26, 27, 28, 29 and 30

of this Standard: items 130, 140, 141, 150 and 151, and 220 to 223) are marked by the symbol "M".

Numeric identifiers of optional data elements (see paragraph 31 of this Standard) are marked by the symbol "O".

<b>Numeric Identifier</b>	<b>Numeric Identifier Description</b>	<b>Mandatory (M) or Optional (O)</b>	<b>Comment</b>
<110>	Applicant name	M	where the name of the applicant is written in characters other than those of the Latin alphabet, the same shall also be indicated in characters of the Latin alphabet either as a mere transliteration or through translation into English
<120>	Title of invention	M	
<130>	File reference	M, in the circumstances specified in paragraph 26 of this Standard	see paragraph 26 of this Standard
<140>	Current patent application	M, in the circumstances specified in paragraph 27 of this Standard	see paragraph 27 of this Standard; the current patent application shall be identified, in the following order, by the two-letter code indicated in accordance with WIPO Standard ST.3 and the application number (in the format used by the industrial property Office with which the current patent application is filed) or, for an international application, by the international application number
<141>	Current filing date	M, in the circumstances specified in paragraph 27 of this Standard	see paragraph 27 of this Standard; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<150>	Earlier patent application	M, in the circumstances specified in paragraph 28 of this Standard	see paragraph 28 of this Standard; the earlier patent application shall be identified, in the following order, by the two-letter code indicated in accordance with WIPO Standard ST.3 and the application number (in the format used by the industrial property Office with which the earlier patent application was filed) or, for an international application, by the international application number
<151>	Earlier application filing date	M, in the circumstances specified in paragraph 28 of this Standard	see paragraph 28 of this Standard; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<160>	Number of SEQ ID NOs	M	
<170>	Software	O	
<210>	Information for SEQ ID NO: x	M	response shall be an integer representing the SEQ ID NO shown
<211>	Length	M	sequence length expressed in number of base pairs or amino acids
<212>	Type	M	type of molecule sequenced in SEQ ID NO: x, either DNA, RNA or PRT; if a nucleotide sequence contains both DNA

			and RNA fragments, the value shall be "DNA"; in addition, the combined DNA/RNA molecule shall be further described in the <220> to <223> feature section
<213>	Organism	M	Genus Species (that is, scientific name) or "Artificial Sequence" or "Unknown"
<220>	Feature	M, in the circumstances specified in paragraph 29 and 30 of this Standard	leave blank; see paragraphs 29 and 30 of this Standard; description of points of biological significance in the sequence in SEQ ID NO: x) (may be repeated depending on the number of features indicated)
<221>	Name/key	M, in the circumstances specified in paragraph 29 of this Standard	see paragraph 29 of this Standard; only those keys as described in Table 5 or 6 of Appendix 2 shall be used
<222>	Location	M, in the circumstances specified in paragraph 29 of this Standard	see paragraph 29 of this Standard; <ul style="list-style-type: none"> <li>- from (number of first base/amino acid in the feature)</li> <li>- to (number of last base/amino acid in the feature)</li> <li>- base pairs (numbers refer to positions of base pairs in a nucleotide sequence)</li> <li>- amino acids (numbers refer to positions of amino acid residues in an amino acid sequence)</li> <li>- whether feature is located on the complementary strand to that filed in the sequence listing</li> </ul>
<223>	Other information	M, in the circumstances specified in paragraphs 29 and 30 of this Standard	see paragraphs 29 and 30 of this Standard; any other relevant information, using language neutral vocabulary, or free text (preferably in English); any free text is to be repeated in the main part of the description in the language thereof (see paragraph 36 of this Standard); where any modified base or modified/unusual L-amino acid appearing in Appendix 2, Tables 2 and 4, is in the sequence, the symbol associated with that base or amino acid from Appendix 2, Tables 2 and 4, should be used
<300>	Publication information	O	leave blank; repeat section for each relevant publication
<301>	Authors	O	
<302>	Title	O	title of publication
<303>	Journal	O	journal name in which data published
<304>	Volume	O	journal volume in which data published
<305>	Issue	O	journal issue number in which data published
<306>	Pages	O	journal page numbers on which data published
<307>	Date	O	journal date on which data published; if possible, the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<308>	Database accession number	O	accession number assigned by database including database name



<309>	Database entry date	O	date of entry in database; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<310>	Document number	O	document number, for patent type citations only; the full document shall specify, in the following order, the two-letter code indicated in accordance with WIPO Standard ST.3, the publication number indicated in accordance with WIPO Standard ST.6, and the kind-of-document code indicated in accordance with WIPO Standard ST.16
<311>	Filing date	O	document filing date, for patent-type citations only; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<312>	Publication date	O	document publication date; for patent-type citations only; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<313>	Relevant residues in SEQ ID NO: x: from to	O	
<400>	Sequence	M	SEQ ID NO: x should follow the numeric identifier and should appear on the line preceding the sequence (see Appendix 3)

## Appendix 2

### Nucleotide and amino acid symbols and feature table

Table 1: List of Nucleotides

Symbol	Meaning	Origin of designation
a	a	<u>a</u> denine
g	g	g <u>u</u> anine
c	c	<u>c</u> ytosine
t	t	<u>t</u> hymine
u	u	<u>u</u> racil
r	g or a	pur <u>r</u> ine
y	t/u or c	pyr <u>y</u> midine
m	a or c	<u>a</u> mino
k	g or t/u	<u>k</u> eto
s	g or c	<u>s</u> trong interactions 3H-bonds
w	a or t/u	<u>w</u> weak interactions 2H-bonds
b	g or c or t/u	not a
d	a or g or t/u	not c
h	a or c or t/u	not g
v	a or g or c	not t, not u
n	a or g or c or t/u, unknown, or other	<u>a</u> ny

Table 2: List of Modified Nucleotides

Symbol	Meaning
ac4c	4-acetylcytidine
chm5u	5-(carboxyhydroxymethyl)uridine
cm	2'-O-methylcytidine
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine
cmnm5u	5-carboxymethylaminomethyluridine
d	dihydrouridine
fm	2'-O-methylpseudouridine
gal q	beta, D-galactosylqueuosine
gm	2'-O-methylguanosine
i	inosine
i6a	N6-isopentenyladenosine
m1a	1-methyladenosine
m1f	1-methylpseudouridine
m1g	1-methylguanosine
m1i	1-methylinosine
m22g	2,2-dimethylguanosine
m2a	2-methyladenosine
m2g	2-methylguanosine
m3c	3-methylcytidine
m5c	5-methylcytidine
m6a	N6-methyladenosine
m7g	7-methylguanosine
mam5u	5-methylaminomethyluridine
mam5s2u	5-methoxyaminomethyl-2-thiouridine
man q	beta, D-mannosylqueuosine
mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
mcm5u	5-methoxycarbonylmethyluridine
mo5u	5-methoxyuridine
ms2i6a	2-methylthio-N6-isopentenyladenosine
ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methylcarbamoyl)threonine
mv	uridine-5-oxyacetic acid-methylester
o5u	uridine-5-oxyacetic acid
osyw	wybutoxosine
p	pseudouridine

q	queuosine
s2c	2-thiocytidine
s2t	5-methyl-2-thiouridine
s2u	2-thiouridine
s4u	4-thiouridine
t	5-methyluridine
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)-carbamoyl)threonine
tm	2'-O-methyl-5-methyluridine
um	2'-O-methyluridine
yw	wybutosine
x	3-(3-amino-3-carboxy-propyl)uridine, (acp3)u

Table 3: List of Amino Acids

Symbol	Meaning
Ala	Alanine
Cys	Cysteine
Asp	Aspartic Acid
Glu	Glutamic Acid
Phe	Phenylalanine
Gly	Glycine
His	Histidine
Ile	Isoleucine
Lys	Lysine
Leu	Leucine
Met	Methionine
Asn	Asparagine
Pro	Proline
Gln	Glutamine
Arg	Arginine
Ser	Serine
Thr	Threonine
Val	Valine
Trp	Tryptophan
Tyr	Tyrosine
Asx	Asp or Asn
Glx	Glu or Gln
Xaa	unknown or other

**Table 4: List of Modified and Unusual Amino Acids**

Symbol	Meaning
Aad	2-Aminoadipic acid
bAad	3-Aminoadipic acid
bAla	beta-Alanine, beta-Aminopropionic acid
Abu	2-Aminobutyric acid
4Abu	4-Aminobutyric acid, piperidinic acid
Acp	6-Aminocaproic acid
Ahe	2-Aminoheptanoic acid
Aib	2-Aminoisobutyric acid
bAib	3-Aminoisobutyric acid
Apm	2-Aminopimelic acid
Dbu	2,4-Diaminobutyric acid
Des	Desmosine
Dpm	2,2'-Diaminopimelic acid
Dpr	2,3-Diaminopropionic acid
EtGly	N-Ethylglycine
EtAsn	N-Ethylasparagine
Hyl	Hydroxylysine
aHyl	allo-Hydroxylysine
3Hyp	3-Hydroxyproline
4Hyp	4-Hydroxyproline
Ide	Isodesmosine
alle	allo-Isoleucine
MeGly	N-Methylglycine, sarcosine
Melle	N-Methylisoleucine
MeLys	6-N-Methyllysine
MeVal	N-Methylvaline
Nva	Norvaline
Nle	Norleucine
Orn	Ornithine

**Table 5: List of Feature Keys Related to Nucleotide Sequences**

Key	Description
allele	a related individual or strain contains stable, alternative forms of the same gene which differs from the presented sequence at this location (and perhaps others)
attenuator	(1) region of DNA at which regulation of termination of transcription occurs, which controls the expression of some bacterial operons; (2) sequence segment located between the promoter and the first structural gene that causes partial termination of transcription
C_region	constant region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; includes one or more exons depending on the particular chain
CAAT_signal	CAAT box; part of a conserved sequence located about 75 bp up-stream of the start point of eukaryotic transcription units which may be involved in RNA polymerase binding; consensus=GG (C or T) CAATCT
CDS	coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon); feature includes amino acid conceptual translation
conflict	independent determinations of the "same" sequence differ at this site or region
D-loop	displacement loop; a region within mitochondrial DNA in which a short stretch of RNA is paired with one strand of DNA, displacing the original partner DNA strand in this region; also used to describe the displacement of a region of one strand of duplex DNA by a single stranded invader in the reaction catalyzed by RecA protein
D-segment	diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain
enhancer	a cis-acting sequence that increases the utilization of (some) eukaryotic promoters, and can function in either orientation and in any location (upstream or downstream) relative to the promoter
exon	region of genome that codes for portion of spliced mRNA; may contain 5'UTR, all CDSs, and 3'UTR
GC_signal	GC box; a conserved GC-rich region located upstream of the start point of eukaryotic transcription units which may occur in multiple copies or in either orientation; consensus=GGGCGG
gene	region of biological interest identified as a gene and for which a name has been assigned
iDNA	intervening DNA; DNA which is eliminated through any of several kinds of recombination
intron	a segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it
J_segment	joining segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains
LTR	long terminal repeat, a sequence directly repeated at both ends of a defined sequence, of the sort typically found in retroviruses
mat_peptide	mature peptide or protein coding sequence; coding sequence for the mature or final peptide or protein product following post-translational modification; the location does not include the stop codon (unlike the corresponding CDS)
misc_binding	site in nucleic acid which covalently or non-covalently binds another moiety that cannot be described by any other Binding key (primer_bind or protein_bind)
misc_difference	feature sequence is different from that presented in the entry and cannot be described by any other Difference key (conflict, unsure, old_sequence, mutation, variation, allele, or modified_base)
misc_feature	region of biological interest which cannot be described by any other feature key; a new or rare feature
misc_recomb	site of any generalized, site-specific or replicative recombination event where there is a breakage and reunion of duplex DNA that cannot be described by other recombination keys (iDNA and virion) or qualifiers of source key (/insertion_seq, /transposon, /proviral)

misc_RNA	any transcript or RNA product that cannot be defined by other RNA keys (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, and snRNA)
misc_signal	any region containing a signal controlling or altering gene function or expression that cannot be described by other Signal keys (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator, and rep_origin)
misc_structure	any secondary or tertiary structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop)
modified_base	the indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value)
mRNA	messenger RNA; includes 5' untranslated region (5'UTR), coding sequences (CDS, exon) and 3' untranslated region (3'UTR)
mutation	a related strain has an abrupt, inheritable change in the sequence at this location
N_region	extra nucleotides inserted between rearranged immunoglobulin segments
old_sequence	the presented sequence revises a previous version of the sequence at this location
polyA_signal	recognition region necessary for endonuclease cleavage of an RNA transcript that is followed by polyadenylation; consensus=AATAAA
polyA_site	site on an RNA transcript to which will be added adenine residues by post-transcriptional polyadenylation
precursor_RNA	any RNA species that is not yet the mature RNA product; may include 5' clipped region (5'clip), 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron), 3' untranslated region (3'UTR), and 3' clipped region (3'clip)
prim_transcript	primary (initial, unprocessed) transcript; includes 5' clipped region (5'clip), 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron), 3' untranslated region (3'UTR), and 3' clipped region (3'clip)
primer_bind	non-covalent primer binding site for initiation of replication, transcription, or reverse transcription; includes site(s) for synthetic, for example, PCR primer elements
promoter	region on a DNA molecule involved in RNA polymerase binding to initiate transcription
protein_bind	non-covalent protein binding site on nucleic acid
RBS	ribosome binding site
repeat_region	region of genome containing repeating units
repeat_unit	single repeat element
rep_origin	origin of replication; starting site for duplication of nucleic acid to give two identical copies
rRNA	mature ribosomal RNA; the RNA component of the ribonucleoprotein particle (ribosome) which assembles amino acids into proteins
S_region	switch region of immunoglobulin heavy chains; involved in the rearrangement of heavy chain DNA leading to the expression of a different immunoglobulin class from the same B-cell
satellite	many tandem repeats (identical or related) of a short basic repeating unit; many have a base composition or other property different from the genome average that allows them to be separated from the bulk (main band) genomic DNA
scRNA	small cytoplasmic RNA; any one of several small cytoplasmic RNA molecules present in the cytoplasm and (sometimes) nucleus of a eukaryote
sig_peptide	signal peptide coding sequence; coding sequence for an N-terminal domain of a secreted protein; this domain is involved in attaching nascent polypeptide to the membrane; leader sequence
snRNA	small nuclear RNA; any one of many small RNA species confined to the nucleus; several of the snRNAs are involved in splicing or other RNA processing reactions

source	identifies the biological source of the specified span of the sequence; this key is mandatory; every entry will have, as a minimum, a single source key spanning the entire sequence; more than one source key per sequence is permissible
stem_loop	hairpin; a double-helical region formed by base-pairing between adjacent (inverted) complementary sequences in a single strand of RNA or DNA
STS	Sequence Tagged Site; short, single-copy DNA sequence that characterizes a mapping landmark on the genome and can be detected by PCR; a region of the genome can be mapped by determining the order of a series of STSs
TATA_signal	TATA box; Goldberg-Hogness box; a conserved AT-rich septamer found about 25 bp before the start point of each eukaryotic RNA polymerase II transcript unit which may be involved in positioning the enzyme for correct initiation; consensus=TATA(A or T)A(A or T)
terminator	sequence of DNA located either at the end of the transcript or adjacent to a promoter region that causes RNA polymerase to terminate transcription; may also be site of binding of repressor protein
transit_peptide	transit peptide coding sequence; coding sequence for an N-terminal domain of a nuclear-encoded organellar protein; this domain is involved in post-translational import of the protein into the organelle
tRNA	mature transfer RNA, a small RNA molecule (75–85 bases long) that mediates the translation of a nucleic acid sequence into an amino acid sequence
unsure	author is unsure of exact sequence in this region
V_region	variable region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for the variable amino terminal portion; can be made up from V_segments, D_segments, N_regions, and J_segments
V_segment	variable segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for most of the variable region (V_region) and the last few amino acids of the leader peptide
variation	a related strain contains stable mutations from the same gene (for example, RFLPs, polymorphisms, etc.) which differ from the presented sequence at this location (and possibly others)
3'clip	3'-most region of a precursor transcript that is clipped off during processing
3'UTR	region at the 3' end of a mature transcript (following the stop codon) that is not translated into a protein
5'clip	5'-most region of a precursor transcript that is clipped off during processing
5'UTR	region at the 5' end of a mature transcript (preceding the initiation codon) that is not translated into a protein
-10_signal	pribnow box; a conserved region about 10 bp upstream of the start point of bacterial transcription units which may be involved in binding RNA polymerase; consensus=TAtAaT
-35_signal	a conserved hexamer about 35 bp upstream of the start point of bacterial transcription units; consensus=TTGACa [ ] or TGTTGACA [ ]



**Table 6: List of Feature Keys Related to Protein Sequences**

Key	Description
CONFLICT	different papers report differing sequences
VARIANT	authors report that sequence variants exist
VARSPLIC	description of sequence variants produced by alternative splicing
MUTAGEN	site which has been experimentally altered
MOD_RES	post-translational modification of a residue
ACETYLATION	N-terminal or other
AMIDATION	generally at the C-terminal of a mature active peptide
BLOCKED	undetermined N- or C-terminal blocking group
FORMYLATION	of the N-terminal methionine
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	of asparagine, aspartic acid, proline or lysine
METHYLATION	generally of lysine or arginine
PHOSPHORYLATION	of serine, threonine, tyrosine, aspartic acid or histidine
PYRROLIDONE CARBOXYLIC ACID	N-terminal glutamate which has formed an internal cyclic lactam
SULFATATION	generally of tyrosine
LIPID	covalent binding of a lipidic moiety
MYRISTATE	myristate group attached through an amide bond to the N-terminal glycine residue of the mature form of a protein or to an internal lysine residue
PALMITATE	palmitate group attached through a thioether bond to a cysteine residue or through an ester bond to a serine or threonine residue
FARNESYL	farnesyl group attached through a thioether bond to a cysteine residue
GERANYL-GERANYL	geranyl-geranyl group attached through a thioether bond to a cysteine residue
GPI-ANCHOR	glycosyl-phosphatidylinositol (GPI) group linked to the alpha-carboxyl group of the C-terminal residue of the mature form of a protein
N-ACYL DIGLYCERIDE	N-terminal cysteine of the mature form of a prokaryotic lipoprotein with an amide-linked fatty acid and a glyceryl group to which two fatty acids are linked by ester linkages
DISULFID	disulfide bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by an intra-chain disulfide bond; if the 'FROM' and 'TO' endpoints are identical, the disulfide bond is an interchain one and the description field indicates the nature of the cross-link
THIOLEST	thiolester bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by the thiolester bond
THIOETH	thioether bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by the thioether bond
CARBOHYD	glycosylation site; the nature of the carbohydrate (if known) is given in the description field
METAL	binding site for a metal ion; the description field indicates the nature of the metal
BINDING	binding site for any chemical group (co-enzyme, prosthetic group, etc.); the chemical nature of the group is given in the description field

SIGNAL	extent of a signal sequence (prepeptide)
TRANSIT	extent of a transit peptide (mitochondrial, chloroplastic, or for a microbody)
PROPEP	extent of a propeptide
CHAIN	extent of a polypeptide chain in the mature protein
PEPTIDE	extent of a released active peptide
DOMAIN	extent of a domain of interest on the sequence; the nature of that domain is given in the description field
CA_BIND	extent of a calcium-binding region
DNA_BIND	extent of a DNA-binding region
NP_BIND	extent of a nucleotide phosphate binding region; the nature of the nucleotide phosphate is indicated in the description field
TRANSMEM	extent of a transmembrane region
ZN_FING	extent of a zinc finger region
SIMILAR	extent of a similarity with another protein sequence; precise information, relative to that sequence is given in the description field
REPEAT	extent of an internal sequence repetition
HELIX	secondary structure: Helices, for example, Alpha-helix, 3(10) helix, or Pi-helix
STRAND	secondary structure: Beta-strand, for example, Hydrogen bonded beta-strand, or Residue in an isolated beta-bridge
TURN	secondary structure Turns, for example, H-bonded turn (3-turn, 4-turn or 5-turn)
ACT_SITE	amino acid(s) involved in the activity of an enzyme
SITE	any other interesting site on the sequence
INIT_MET	the sequence is known to start with an initiator methionine
NON_TER	the residue at an extremity of the sequence is not the terminal residue; if applied to position 1, this signifies that the first position is not the N-terminus of the complete molecule; if applied to the last position, it signifies that this position is not the C-terminus of the complete molecule; there is no description field for this key
NON_CONS	non consecutive residues; indicates that two residues in a sequence are not consecutive and that there are a number of unsequenced residues between them
UNSURE	uncertainties in the sequence; used to describe region(s) of a sequence for which the authors are unsure about the sequence assignment

## Appendix 3

### Specimen sequence listing

```

<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  cggccctct  cgcgctcctc  tcgcgcctct  ctctcgctct  cctctcgctc          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
    
```

<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

## 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
<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
<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 ; <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>- paires de bases (numéros des positions des paires de 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 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)
<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)

## 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	<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



Tableau 2 : liste des nucléotides modifiés

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-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
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

Tableau 3 : liste des acides aminés

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

Tableau 4 : liste des acides aminés modifiés ou peu connus

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
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

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 ("displacement loop") ; 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 ("diversity segment") 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	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 segment") 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 ("long terminal repeat") ; 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 répliquative, 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, scrRNA 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
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
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
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 auteurs signalent qu'il existe des variants de la séquence
VARSP LIC	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 glycérite 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 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)
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



### 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  cggccctct  cgcgctcctc  tcgcgcctct  ctctcgctct  cctctcgctc          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
    
```

<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

## Mitteilung des Europäischen Patentamts vom 2. Oktober über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen

Für internationale Anmeldungen nach dem PCT gilt ab 1. Juli 1998 der "Standard über die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Anmeldungen nach dem PCT"<sup>1</sup> (abgekürzt "PCT-Standard", siehe Regeln 5.2, 12.1 (d) und 13<sup>ter</sup> PCT sowie Abschnitt 208 und Anhang C der PCT-Verwaltungsvorschriften). Zum selben Zeitpunkt ist WIPO-Standard ST. 25 für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in Patentanmeldungen in Kraft getreten. Dieser neue WIPO-Standard, der die bisherigen WIPO-Standards ST. 23 und ST. 24 ersetzt, empfiehlt den Patentämtern, den PCT-Standard entsprechend auf alle Patentanmeldungen, bei denen es sich nicht um internationale Anmeldungen handelt, anzuwenden.

Gemäß Beschluß des Präsidenten des EPA vom 2. Oktober 1998<sup>2</sup> (nachfolgend "Beschluß des Präsidenten" genannt) ist der WIPO-Standard ST. 25 und damit der PCT-Standard für das europäische Patenterteilungsverfahren ab dem 1. Januar 1999 maßgebend.

Diese Mitteilung ergänzt den Beschluß des Präsidenten und erläutert das Verfahren nach dem EPÜ und dem PCT.

### I. Allgemeines

Seit dem 1. Januar 1993 ist in Verfahren vor dem EPA nach dem EPÜ und dem PCT die Darstellung von Nucleotid- und Aminosäuresequenzen in Form eines standardisierten Sequenzprotokolls sowie dessen Einreichung auf einem elektronischen Datenträger zwingend vorgeschrieben. Der neue WIPO Standard ST. 25 ersetzt die bisherigen Standards ST. 23 und ST. 24. Damit werden neben der Berücksichtigung des technologischen Fortschritts Vereinfachungen, Klarstellungen und Erleichterungen für die Anmelder

<sup>1</sup> Abgedruckt in der PCT-Gazette (Sonderausgabe) vom 25.6.1998.

<sup>2</sup> Abgedruckt in dieser Beilage auf Seite 1.

## Notice from the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings

The "Standard for the presentation of nucleotide and amino acid sequence listings in international applications under the PCT"<sup>1</sup> (or "PCT Standard", see Rules 5.2, 12.1(d) and 13<sup>ter</sup> PCT as well as Section 208 of and Annex C to the PCT Administrative Instructions) came into force for international applications under the PCT on 1 July 1998. WIPO Standard ST.25 for the presentation of nucleotide and amino acid sequences in patent applications also came into force on that date. This new WIPO Standard, which replaces WIPO Standards ST.23 and ST.24, recommends that patent offices also apply the PCT Standard mutatis mutandis to all patent applications in addition to international applications.

According to the decision of the President of the European Patent Office dated 2 October 1998 (hereinafter referred to as the "decision of the President"), WIPO Standard ST.25, and hence the PCT Standard, will enter into force for the European patent grant procedure on 1 January 1999.

This notice supplements the decision of the President and is intended to clarify the EPC and PCT procedures.

### I. General remarks

The presentation of nucleotide and amino acid sequences in the form of a standardised sequence listing as well as the filing thereof on an electronic data carrier has been mandatory in proceedings before the EPO under the EPC and the PCT since 1 January 1993. The new WIPO Standard ST.25 replaces Standards ST.23 and ST.24. It not only takes account of advances in technology, but also simplifies, clarifies and facilitates procedures for the applicant. In particular, it will no longer be necessary to translate a sequence

<sup>1</sup> See special edition of the PCT Gazette of 25 June 1998.

<sup>2</sup> See page 1 of this supplement.

## Communiqué de l'Office européen des brevets en date du 2 octobre 1998, relatif à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences

A compter du 1<sup>er</sup> juillet 1998, la norme applicable aux demandes internationales au titre du PCT est 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 PCT"<sup>1</sup> (en abrégé "norme du PCT", cf. règles 5.2, 12.1 d) et 13<sup>ter</sup> PCT ainsi que l'instruction administrative 208 et l'annexe C des Instructions administratives du PCT). A la même date est entrée en vigueur la norme ST. 25 de l'OMPI relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet. Cette nouvelle norme de l'OMPI, qui remplace les normes ST. 23 et ST. 24 de l'OMPI qui étaient applicables jusqu'ici, recommande aux offices de brevets d'appliquer par analogie la norme du PCT pour toutes les demandes de brevet autres que les demandes internationales.

Conformément à la décision du Président de l'OEB en date du 2 octobre 1998<sup>2</sup> (dénommée ci-après "décision du Président"), la norme ST. 25 de l'OMPI (et donc la norme du PCT) est applicable à compter du 1<sup>er</sup> janvier 1999 pour la procédure de délivrance de brevets européens.

Le présent communiqué complète la décision du Président et fournit des explications concernant la procédure selon la CBE et selon le PCT.

### I. Généralités

Dans les procédures au titre de la CBE ou du PCT se déroulant devant l'OEB, il est exigé impérativement depuis le 1<sup>er</sup> janvier 1993 la présentation d'un listage des séquences de nucléotides et d'acides aminés sous forme de listage des séquences normalisé, et le dépôt de ce listage sur un support électronique de données. La nouvelle norme ST. 25 de l'OMPI remplace les normes ST. 23 et ST. 24 qui s'appliquaient jusque là. Il est tenu compte ainsi du progrès technique, et de plus la situation se voit simplifiée, clarifiée et facilitée pour

<sup>1</sup> Reproduite dans la Gazette du PCT, numéro spécial du 25 juin 1998.

<sup>2</sup> Reproduite dans le présent supplément, page 1.

eingeführt. Insbesondere soll der Zwang zur Übersetzung eines Sequenzprotokolls bei Verwendung für verschiedene Anmeldungen in verschiedenen Sprachen dadurch entfallen, daß das Sequenzprotokoll unter Verzicht auf unnötige sprachliche Angaben (ersetzt durch numerische Codes) und unter Verwendung eines kontrollierten, von der Fachwelt allgemein anerkannten Vokabulars formuliert wird.

## II. Europäisches Patenterteilungsverfahren

### 1. Sequenzprotokoll

1.1 Die Darstellung von Nucleotid- und Aminosäuresequenzen im vorgeschriebenen Sequenzprotokoll muß dem im Anhang zum Beschluß des Präsidenten wiedergegebenen WIPO-Standard ST. 25 und damit dem PCT-Standard entsprechen (im folgenden "Standard" genannt).

1.2 Ein Sequenzprotokoll ist nur für solche Sequenzen einzureichen, die der Definition laut Nummer 2 (ii) des Standards entsprechen und sich auf die in der Anmeldung offenbarte Erfindung beziehen.

Soweit in der Beschreibung auf Sequenzen Bezug genommen wird, die zum Stand der Technik gehören und in allgemein zugänglichen Datenbanken auffindbar sind, ist es nicht erforderlich, solche Sequenzen in das Sequenzprotokoll aufzunehmen. Jedoch ist jede solche Sequenz hinreichend durch Angabe ihrer Datenbank-Zugangsnummer zu identifizieren.

1.3 Wird ein Sequenzprotokoll zusammen mit den Anmeldungsunterlagen, d. h. als Bestandteil der Beschreibung eingereicht, so sollen die Sequenzen nicht in der Beschreibung wiederholt werden (siehe Standard, Nummer 3).

1.4 Das Sequenzprotokoll kann entweder in der Sprache der Anmeldung oder (bevorzugt) in Englisch abgefaßt werden. Sind Sequenzprotokoll einerseits und Beschreibung und Ansprüche andererseits nicht in derselben Sprache abgefaßt und enthält das Sequenzprotokoll "freien Text" im Sinne des Standards (Nummer 33), so wird empfohlen, den freien Text am Ende der Beschreibung in der Sprache der Beschreibung gemäß Nummer 36 des Standards wiederzugeben (siehe Beispiele in der Anlage zu dieser Mitteilung). Damit kann vermieden werden, daß Merkmale, die nur im sogenannten "freien Text" des Sequenzprotokolls angegeben sind, in einzelnen Ver-

listing used for different applications in different languages, as sequence listings can be formulated without unnecessary text (replaced by numerical codes) and using a limited vocabulary of terms generally recognised in the field.

## II. European patent grant procedure

### 1. Sequence listing

1.1 The presentation of nucleotide and amino acid sequences in the prescribed sequence listing must comply with WIPO Standard ST.25, reproduced in the annex to the decision of the President, and thus with the PCT Standard (hereinafter referred to as the "Standard").

1.2 A sequence listing need only be filed for those sequences matching the definition given in the Standard, paragraph 2 (ii), and relating to the invention disclosed in the application.

Where the description refers to sequences which belong to the prior art and can be found in generally accessible databases, it will not be necessary to include such sequences in the sequence listing. However, each such sequence must be identified adequately by its database access number.

1.3 Where a sequence listing is filed together with the application documents, ie as part of the description, the sequences should not be repeated in the description (see Standard, paragraph 3).

1.4 The sequence listing may be written either in the language of filing or (preferably) in English. Where the sequence listing on the one hand and the description and claims on the other are not in the same language, and the sequence listing contains "free text" within the meaning of the Standard (paragraph 33), it is recommended that the free text be repeated at the end of the description in the language of the description in accordance with paragraph 36 of the Standard (see examples in the annex to this notice). This will ensure that features which are given only in the so-called "free text" of the sequence listing are not deemed in individual contracting states to the EPC not to

les demandeurs. Il n'y aura notamment plus lieu d'exiger la traduction d'un même listage des séquences figurant dans différentes demandes rédigées dans des langues différentes. Du fait que les indications linguistiques superflues disparaîtront du listage des séquences (elles seront remplacées par des codes numériques) et qu'il sera utilisé un vocabulaire contrôlé, universellement reconnu par les hommes du métier.

## II. Procédure de délivrance des brevets européens

### 1. Listage des séquences

1.1 La présentation du listage prescrit des séquences de nucléotides et d'acides aminés doit être conforme à la norme ST. 25 de l'OMPI reproduite dans l'annexe à la décision du Président, et donc à la norme du PCT (dénommée ci-après "la norme").

1.2 Il n'y a lieu de déposer de listage des séquences que pour les séquences correspondant à la définition figurant au paragraphe 2(ii) de la norme et se rapportant à l'invention exposée dans la demande.

Dans la mesure où il est fait référence dans la description à des séquences qui sont comprises dans l'état de la technique et que l'on peut trouver dans des banques de données accessibles au public, il n'est pas nécessaire d'intégrer ces séquences dans le listage des séquences. Toutefois il convient d'identifier suffisamment toutes ces séquences en indiquant leur numéro d'accès dans la banque de données.

1.3 Si un listage des séquences est déposé en même temps que les pièces de la demande, c'est-à-dire en tant que partie de la description, il n'y a pas lieu de reprendre les séquences dans la description (cf. norme, paragraphe 3).

1.4 Le listage des séquences peut être rédigé soit dans la langue de la demande, soit (de préférence) en anglais. Si le listage des séquences, d'une part, et la description et les revendications, d'autre part, ne sont pas rédigés dans la même langue, et que le listage des séquences ne comporte pas de "texte libre" au sens de la norme (paragraphe 33), il est recommandé de répéter le texte libre à la fin de la description, dans la langue de la description, conformément au paragraphe 36 de la norme, ce qui permettra d'éviter que dans certains Etats parties à la CBE des caractéristiques qui ne sont indiquées que dans le "texte libre" du listage des séquences ne soient pas

tragsstaaten des EPÜ aufgrund der Sprache nicht als Gegenstand der ursprünglichen Offenbarung angesehen werden.

## 2. Elektronischer Datenträger

2.1 Die Einreichung des Sequenzprotokolls auf einem elektronischen Datenträger zusätzlich zur schriftlichen Form ist vorgeschrieben. Die näheren Einzelheiten über die zugelassenen Datenträger, ihre Kennzeichnung, das Format der darauf gespeicherten Information und die erforderliche Hardware sind dem Standard (Nummern 39 ff.) zu entnehmen.

2.2 Das EPA empfiehlt die Verwendung der Software **Patentin**, die überarbeitet worden ist (Version 2.0) und den Anmeldern kostenlos zur Verfügung steht. Durch Verwendung dieser Software (in der jeweils gültigen Version) kann der Anmelder erreichen, daß die Formerfordernisse nach Regel 27a (1) EPÜ erfüllt sind. Die Software kann von der Website des EPA heruntergeladen werden und eignet sich für PCs, die mit dem Betriebssystem Windows (Version 3.1 oder höher) ausgestattet sind. Weitere Informationen sind erhältlich bei:

Europäisches Patentamt  
Programm STRAND  
Patentlaan 2  
NL-2288 EE Rijswijk  
Fax: (+31-70) 340 3992  
E-mail: Patentin\_help.europatent.org

### 2.3 Beschädigter oder unlesbarer Datenträger

Ist der zusammen mit der Anmeldung eingereichte Datenträger beschädigt oder aus einem anderen Grund nicht lesbar, so wird der Anmelder aufgefordert, einen einwandfreien Datenträger innerhalb einer nicht verlängerbaren Frist von einem Monat ab Zustellung der entsprechenden Mitteilung nachzureichen.

### 3. Erklärung gemäß Regel 27a (2) EPÜ

Der vorgeschriebene Datenträger wird für die EDV-gestützte europäische oder internationale Recherche sowie, nach der Veröffentlichung der Anmeldung, für die laufende Aktualisierung der internationalen Patent-Sequenzdatenbank benötigt.

Die auf dem Datenträger gespeicherte Information muß mit dem schriftlichen Sequenzprotokoll über-

be the subject-matter of the original disclosure because of the language used.

## 2. Electronic data carriers

2.1 Sequence listings must be submitted on an electronic data carrier as well as in written form. Further details with regard to authorised data carriers, their identification, the format for the information recorded on them and the necessary hardware are set out in the Standard (paragraphs 39 ff).

2.2 The EPO recommends the use of **Patentin** software, which has been upgraded (version 2.0) and is available to applicants free of charge. By using the (applicable) version of this software, applicants can ensure that the physical requirements under Rule 27a(1) EPC are met. The software can be downloaded from the EPO's website and is suitable for use with Windows 3.1 or above. Further information can be obtained from:

European Patent Office  
STRAND program  
Patentlaan 2  
NL-2288 EE Rijswijk  
Fax: (+31-70) 340 3992  
e-mail: Patentin\_help.europatent.org

### 2.3 Damaged or unreadable data carriers

If a data carrier submitted to the EPO is damaged, or unreadable for some other reason, the applicant will be invited to submit a flawless replacement within a non-extendable period of one month from notification of the communication.

### 3. Statement under Rule 27a(2) EPC

The prescribed data carrier is required for EDP-based European or international searches and, after publication of the application, for continuous updating of the international patent sequence database.

The information recorded on the data carrier must be identical to the written sequence listing. This must

considérées comme l'objet de la divulgation initiale, en raison de la différence existant entre la langue utilisée dans la description et celle utilisée dans le texte libre.

## 2. Support électronique de données

2.1 La production du listage des séquences sur un support électronique de données en plus du listage imprimé est obligatoire. Des précisions concernant les supports électroniques admis, leur identification, le format des informations y figurant et le matériel nécessaire sont données dans la norme (paragraphes 39 s.)

2.2 L'OEB recommande l'utilisation du logiciel **Patentin**, qui a été revu (version 2.0) et est mis gratuitement à la disposition des demandeurs. Le demandeur qui utilise ce logiciel (dans la version en vigueur) est en mesure de satisfaire aux conditions de forme prescrites à la règle 27bis (1) CBE. Le logiciel peut être téléchargé à partir du site web de l'OEB et convient pour les PC qui sont équipés du programme d'exploitation Windows (version 3.1 ou versions suivantes). Pour plus d'informations s'adresser à :

Office européen des brevets  
Programme STRAND  
Patentlaan 2  
NL-2288 EE Rijswijk  
Fax: (+31-70) 340 3992  
E-mail : Patentin\_help.europatent.org

### 2.3 Support de données endommagé ou illisible

Si le support de données déposé en même temps que la demande est endommagé ou illisible pour quelque autre raison, le demandeur est invité à déposer un support de données en parfait état dans un délai d'un mois à compter de la signification de la notification correspondante ; ce délai ne peut être prolongé.

### 3. Déclaration prévue à la règle 27bis (2) CBE

Le support de données prescrit est nécessaire pour la recherche européenne ou internationale assistée par ordinateur ainsi que, après la publication de la demande, pour la mise à jour constante de la banque internationale de données relatives aux séquences contenues dans les brevets.

L'information figurant sur ce support doit être identique à celle que contient le listage imprimé des séquen-

einstimmen. Der Anmelder oder sein Vertreter hat dies sicherzustellen und eine entsprechende Erklärung gemäß Regel 27a (2) EPÜ zusammen mit dem Datenträger einzureichen (siehe auch Standard, Nummer 40). Eine solche Erklärung ist im Erteilungsantrag enthalten (EPA Form 1001, Feld 31)<sup>3</sup>. Sie kann als Mustererklärung auch bei Nachreichung des Datenträgers verwendet werden.

Stellt die Prüfungsabteilung fest, daß der eingereichte Datenträger das Identitätserfordernis nicht erfüllt und somit die Erklärung gemäß Regel 27a (2) EPÜ unrichtig ist, so kann dies zur Zurückweisung der Anmeldung nach Artikel 97 (1) EPÜ führen. Darüber hinaus kann ein anderer Stand der Technik einschlägig sein, als er der Prüfung zugrunde gelegt wurde. Dies kann zu einer anderen Beurteilung der Patentierbarkeitsvoraussetzungen durch die jeweils zuständigen Instanzen auch nach Patenterteilung führen.

#### 4. Nachreichung eines Sequenzprotokolls

4.1 Ist ein Sequenzprotokoll einzureichen, so ist es den Anmeldeunterlagen als Teil der Beschreibung beizufügen (siehe Standard, Nummer 3).

4.2 Gleichwohl nimmt das EPA auch nach dem Anmeldetag eingereichte Sequenzprotokolle entgegen, sofern der vorgeschriebene Datenträger und die Erklärungen gemäß Regel 27a (2) und (3) EPÜ vorliegen, weil dies für die Durchführung der Recherche und die Fortschreibung der internationalen Patent-Sequenzdatenbank sinnvoll ist.

4.3 Ein später eingereichtes Sequenzprotokoll ist jedoch nicht Bestandteil der Beschreibung und kann daher für die Bestimmung des ursprünglichen Offenbarungsgehalts der Anmeldung nicht herangezogen werden (Regel 27a (4) EPÜ).

4.4 Ist der Anmelder nicht imstande, das Sequenzprotokoll mit der Anmeldung einzureichen, sollte er es von sich aus baldmöglichst nachreichen und die weiteren Erfordernisse nach Regel 27a (2) und (3) EPÜ erfüllen (Einreichung des vorgeschriebenen Datenträgers und der vorgesehenen Erklärungen, siehe auch Punkt II.5). Andernfalls fordert ihn die Eingangsstelle auf, die festgestellten Mängel innerhalb einer **Frist von zwei Mona-**

be ensured by the applicant or his representative, who must file together with the data carrier a statement in accordance with Rule 27a(2) EPC (see also Standard, paragraph 40). The statement is contained in the Request for Grant (EPO Form 1001, section 31)<sup>3</sup>, and may also be used as a model statement when the data carrier is filed subsequently.

If the examining division finds that the data carrier submitted does not comply with the requirement that it be identical and the statement under Rule 27a (2) EPC is therefore incorrect, this may lead to refusal of the application under Article 97(1) EPC. Furthermore, a different prior art than that on which the examination was based may be relevant. This may lead to a different assessment by the relevant departments or courts of the conditions for patentability, even after a patent has been granted.

#### 4. Subsequent filing of a sequence listing

4.1 If a sequence listing has to be filed, it should be annexed to the application documents as part of the description (see Standard, paragraph 3).

4.2 Nevertheless, the EPO accepts sequence listings filed after the date of filing so long as the prescribed data carrier and the statements under Rule 27a(2) and (3) EPC have been received because they can be useful in the search and for updating the international patent sequence database.

4.3 A sequence listing filed later does not, however, form part of the description and cannot therefore be used in determining the original disclosed content of the application (Rule 27a(4) EPC).

4.4 If the applicant is unable to file the sequence listing with the description, he should file it without prompting as soon as possible thereafter and comply with the other requirements under Rule 27a(2) and (3) EPC (see also point II.5 for submission of the prescribed data carrier and statements). Otherwise, the Receiving Section will invite him to remedy the deficiencies within a **period of two months** (Article 91(2),

ces. Le demandeur ou son mandataire est tenu de prendre les mesures nécessaires à cet effet et de joindre au support de données une déclaration en ce sens, conformément à la règle 27bis (2) CBE (cf. également la norme, paragraphe 40). Une telle déclaration figure dans la requête en délivrance (OEB Form 1001, rubrique 31)<sup>3</sup>. Elle peut également être utilisée comme déclaration type en cas de production ultérieure du support de données.

Si la division d'examen constate que le support de données ne satisfait pas à l'exigence d'identité avec la liste écrite et que la déclaration selon la règle 27bis (2) CBE est donc inexacte, la demande peut être rejetée conformément à l'article 97(1) CBE. Il se peut en outre qu'un état de la technique autre que celui qui a été pris en compte pour l'examen soit pertinent et qu'il en résulte une appréciation divergente par les instances compétentes des conditions de brevetabilité même après la délivrance du brevet.

#### 4. Dépôt ultérieur d'un listage des séquences

4.1 Tout listage des séquences doit être joint aux pièces de la demande comme faisant partie de la description (cf. norme, paragraphe 3).

4.2 L'OEB accepte néanmoins les listages déposés après la date de dépôt, dans la mesure où le support de données prescrit et les déclarations selon la règle 27bis (2) et (3) CBE ont été produits, étant donné que ces informations sont utiles pour effectuer la recherche et actualiser la banque internationale de données relatives aux séquences contenues dans les brevets.

4.3 Un listage de séquences déposé ultérieurement ne fait toutefois pas partie de la description et ne peut donc servir à déterminer le contenu initial de la demande (règle 27bis (4) CBE).

4.4 Si le demandeur n'est pas en mesure de déposer le listage des séquences avec la demande, il a intérêt à en effectuer spontanément le dépôt ultérieurement dans les meilleurs délais, en veillant à remplir les autres conditions énoncées à la règle 27bis (2) et (3) CBE (production du support de données prescrit et des déclarations prévues; cf. également point II.5). Sinon, la section de dépôt l'invite à remédier dans un

<sup>3</sup> Abgedruckt im ABI. EPA 1997, 251.

<sup>3</sup> Published in OJ EPO 1997, 251.

<sup>3</sup> Reproduit au JO OEB 1997, 251.

ten zu beheben (Art. 91 (2), Regel 41 (1) EPÜ und Artikel 3 des Beschlusses des Präsidenten). Die Frist wird **einmal um höchstens zwei Monate** verlängert, sofern der Anmelder dies vor Ablauf der Frist beantragt. In begründeten Ausnahmefällen kann die Frist auch von Amts wegen verlängert werden, wenn nicht alle Mängel innerhalb der gesetzten Frist beseitigt worden sind. Wird innerhalb dieser Frist ein **beschädigter oder aus einem anderen Grund unlesbarer Datenträger** eingereicht, so wird dem Anmelder eine besondere, nicht verlängerbare Frist von einem Monat ab Zustellung der entsprechenden Mitteilung zur Nachreichung eines einwandfreien Datenträgers eingeräumt.

##### 5. Erklärung gemäß Regel 27a (3) EPÜ

Ein nachgereichtes Sequenzprotokoll darf nichts enthalten, das über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht, da die Recherche auf der Grundlage des nachgereichten Sequenzprotokolls durchgeführt wird. Das nachgereichte Sequenzprotokoll darf daher nur diejenige Sequenzinformation – in standardisierter Form – enthalten, die in den ursprünglichen Anmeldungsunterlagen bereits vorhanden war. Es ist unbedingt darauf zu achten, daß das nachgereichte Sequenzprotokoll nicht mehr und nicht weniger Sequenzen als die in der ursprünglichen Beschreibung offenbarten Sequenzen enthält und die Numerierung dieser Sequenzen unverändert bleibt. Der Anmelder hat dies sicherzustellen und eine entsprechende Erklärung gemäß Regel 27a (3) EPÜ einzureichen. Es wird empfohlen, hierfür die in der Fußnote abgedruckte Mustererklärung<sup>4</sup> zu verwenden. Für den ebenfalls einzureichenden Datenträger gilt Punkt II.3.

##### 6. Rechtsverlust

Werden die Erfordernisse der Regel 27a EPÜ in Verbindung mit dem Beschluß des Präsidenten nicht rechtzeitig, gegebenenfalls nach Aufforderung der Eingangsstelle gemäß Regel 41 (1) EPÜ, erfüllt, so wird die Anmeldung zurückgewiesen (Art. 91 (1) b) und (3) EPÜ).

Der Anmelder kann die Weiterbehandlung der Anmeldung nach Artikel 121 EPÜ beantragen.

Rule 41(1) EPC and Article 3 of the decision of the President). This period may be extended **once only for a maximum of two months**, provided that the applicant requests such an extension before the period expires. In exceptional, substantiated cases, the period may be extended by the Receiving Section of its own motion where not all the deficiencies have been remedied within the prescribed period. If a **data carrier is filed which is damaged, or unreadable for some other reason**, the applicant will be granted a special, non-extendable period of one month from notification of the communication within which to file a flawless replacement.

##### 5. Statement under Rule 27a(3) EPC

A subsequently filed sequence listing may not include any matter which goes beyond the content of the application as filed, since the search is to be carried out on the basis of the subsequently filed sequence listing. The subsequently filed sequence listing may therefore contain only the sequence information – in a standardised form – that was already in the original application documents. The subsequently filed sequence listing should contain no more and no fewer sequences than disclosed in the original description, and their numbering should remain unchanged. This must be ensured by the applicant, who must file a statement to that effect in accordance with Rule 27a(3) EPC. It is recommended that the specimen statement in footnote<sup>4</sup> below be used for this purpose. Point II.3 is applicable for the data carrier which must also be submitted.

##### 6. Loss of rights

If the requirements laid down in Rule 27a EPC in conjunction with the decision of the President are not met in good time – if applicable, after an invitation from the Receiving Section under Rule 41(1) EPC – the application is refused (Art. 91(1)(b) and (3) EPC).

The applicant may request further processing of the application under Article 121 EPC.

**délai de deux mois** aux irrégularités constatées (art. 91(2), règle 41(1) CBE et art. 3 de la décision du Président). Le délai est prolongé **une seule fois de deux mois au maximum**, si le demandeur présente une requête en ce sens avant l'expiration du délai. Dans des cas exceptionnels, si la situation le justifie, le délai peut également être prolongé d'office, s'il n'a pas été remédié à toutes les irrégularités dans le délai fixé. S'il est déposé dans ce délai un **support de données endommagé ou rendu illisible pour quelque autre raison**, il est imparti au demandeur un délai spécial non prorogeable d'un mois à compter de la signification de la notification correspondante invitant le demandeur à déposer un support de données en parfait état.

##### 5. Déclaration prévue à la règle 27bis (3) CBE

Un listage des séquences déposé ultérieurement ne doit contenir aucun élément s'étendant au-delà du contenu de la demande telle qu'elle a été déposée, puisque la recherche est exécutée sur la base de ce listage. Le listage des séquences déposé ultérieurement ne doit contenir – sous forme normalisée – que les informations concernant les séquences qui figuraient déjà dans les pièces initiales de la demande. Il faut absolument veiller à ce que le listage des séquences déposé ultérieurement ne contienne ni plus ni moins de séquences que celles qui sont divulguées dans la description initiale et à ce que la numérotation de ces séquences demeure inchangée. Le demandeur est tenu de faire le nécessaire à cet effet et de produire une déclaration en ce sens conformément à la règle 27bis (3) CBE. Il est recommandé d'utiliser à cette fin la déclaration type reproduite dans la note en bas de page<sup>4</sup>. Le point II.3 est applicable au support de données que le demandeur doit en outre produire.

##### 6. Perte de droits

Si le demandeur ne satisfait pas en temps utile aux conditions énoncées à la règle 27bis CBE en liaison avec la décision du Président de l'OEB, le cas échéant après y avoir été invité par la section de dépôt conformément à la règle 41 (1) CBE, la demande est rejetée (art. 91(1)(b) et (3) CBE).

Le demandeur peut demander la poursuite de la procédure conformément à l'article 121 CBE.

<sup>4</sup> "Es wird hiermit erklärt, daß das Sequenzprotokoll nicht über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht."

<sup>4</sup> "It is hereby declared that the sequence listing does not go beyond the content of the application as filed."

<sup>4</sup> "Il est déclaré par la présente que le listage des séquences ne s'étend pas au-delà du contenu de la demande telle qu'elle a été déposée."

### III. Verfahren vor dem EPA als Internationaler Recherchenbehörde oder als mit der internationalen vorläufigen Prüfung beauftragter Behörde

1. Sequenzprotokolle in maschinenlesbarer Form werden für die Durchführung der internationalen Recherche ebenso benötigt wie für die europäische Recherche. Sie müssen dem PCT-Standard entsprechen (Regel 5.2 PCT sowie Abschnitt 208 und Anhang C der PCT-Verwaltungsvorschriften).

Liegt ein diesem Standard entsprechendes Sequenzprotokoll zu Beginn der internationalen Recherche nicht oder nicht auf elektronischem Datenträger vor, so wird der Anmelder vom EPA aufgefordert, das Sequenzprotokoll und/oder den vorgeschriebenen Datenträger innerhalb einer **nicht verlängerbaren Frist von maximal zwei Monaten** nachzureichen (Regel 13<sup>ter</sup>.1 a) PCT). Diese wird wegen der engen Fristen der Regel 42 PCT für die Erstellung des Recherchenberichts regelmäßig nur **einen Monat** betragen.

2. Bei Nachreichung des Sequenzprotokolls und/oder des Datenträgers hat der Anmelder zu erklären, daß das nachgereichte Sequenzprotokoll nicht über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht und/oder die auf dem Datenträger gespeicherte Information mit dem schriftlichen Sequenzprotokoll übereinstimmt (Art. 4 des Beschlusses des Präsidenten und Standard, Nummern 37 und 40).

3. Werden diese Erfordernisse nicht oder nicht rechtzeitig erfüllt, so führt das EPA keine internationale Recherche durch, soweit eine sinnvolle Recherche nicht möglich ist (Regel 13<sup>ter</sup>.1 c) PCT). In diesem Fall wird der internationale Recherchenbericht ganz oder teilweise durch die Erklärung gemäß Artikel 17 (2) PCT ersetzt.

4. Ist der internationale Recherchenbericht aus den vorgenannten Gründen nicht oder nur teilweise erstellt worden, so führt das EPA als mit der internationalen vorläufigen Prüfung beauftragte Behörde keine internationale vorläufige Prüfung durch, soweit eine sinnvolle Prüfung nicht möglich ist (Regel 66.1 e) PCT). Dies wird dem Anmelder mitgeteilt und im internationalen vorläufigen Prüfungsbericht vermerkt (Regel 66.2 a) vii)

### III. Proceedings before the EPO as International Searching Authority or International Preliminary Examining Authority

1. Sequence listings in machine readable form are just as necessary in carrying out the international search as they are for the European search. They must comply with the PCT Standard (Rule 5.2 PCT as well as Section 208 of and Annex C to the PCT Administrative Instructions).

If at the start of the international search a sequence listing complying with this Standard has not been received or has not been received on an electronic data carrier, the EPO invites the applicant to file the sequence listing and/or the prescribed data carrier within a **non-extendable period of not more than two months** (Rule 13<sup>ter</sup>.1(a) PCT). Because of the short time limits in Rule 42 PCT for the establishment of the search report, this period will normally be only **one month**.

2. If the sequence listing and/or data carrier is filed subsequently, the applicant must state that the sequence listing so filed does not include matter which goes beyond the content of the application as filed and/or that the information recorded on the data carrier is identical to the written sequence listing (Article 4 of the decision of the President, and Standard, paragraphs 37 and 40).

3. If these requirements are not met or not met in due time, the EPO does not perform the international search where a meaningful search cannot be carried out (Rule 13<sup>ter</sup>.1(c) PCT). In this case the international search report is replaced in full or in part by the statement under Article 17(2) PCT.

4. If the international search report has not been drawn up in full or in part for the aforementioned reasons, the EPO as International Preliminary Examining Authority does not perform the international preliminary examination where a meaningful examination cannot be carried out (Rule 66.1(e) PCT). This is notified to the applicant and stated in the international preliminary examination report (Rules 66.2(a)(vii) and 70.12(iv))

### III. Procédure devant l'OEB agissant en qualité d'administration chargée de la recherche internationale ou d'administration chargée de l'examen préliminaire international

1. Les listages des séquences sous forme déchiffirable par machine sont nécessaires tant pour la recherche internationale que pour la recherche européenne. Ils doivent être établis conformément à la norme du PCT (règle 5.2 PCT ainsi que instruction administrative 208 et annexe C des Instructions administratives du PCT).

Si un listage des séquences conforme à cette norme n'est pas parvenu à l'OEB ou bien s'il n'est pas disponible sur support électronique de données lorsque la recherche internationale est entreprise, le demandeur est invité par l'OEB à fournir **dans un délai non prorogable de deux mois au maximum** le listage et/ou le support de données prescrit (règle 13<sup>ter</sup>.1 a) PCT). Ce délai ne sera normalement que d'**un mois** en raison des délais très stricts fixés par la règle 42 PCT pour l'établissement du rapport de recherche internationale.

2. Lorsqu'il fournit ultérieurement le listage des séquences et/ou le support de données, le demandeur est tenu de déclarer que ce listage ne s'étend pas au-delà du contenu de la demande telle qu'elle a été déposée et/ou que l'information figurant sur ce support est identique à celle que contient le listage imprimé (art. 4 de la décision du Président et norme, paragraphes 37 et 40).

3. Si ces conditions ne sont pas remplies ou ne le sont pas dans les délais, l'OEB ne procède pas à la recherche internationale dans la mesure où une recherche significative ne peut pas être effectuée (règle 13<sup>ter</sup>.1 c) PCT). Dans ce cas, le rapport de recherche internationale est remplacé en tout ou en partie par la déclaration prévue à l'article 17(2) PCT.

4. Si le rapport de recherche internationale n'a pas été établi, ou n'a été établi qu'en partie, pour les raisons qui viennent d'être énoncées, l'OEB agissant en qualité d'administration chargée de l'examen préliminaire international ne procède pas à l'examen préliminaire international dans la mesure où un examen significatif ne peut pas être effectué (règle 66.1 e) PCT). L'OEB en informe le demandeur et l'indique dans le rapport



und 70.12 iv) PCT). Eine Aufforderung zur Einreichung eines Sequenzprotokolls gemäß Regel 13<sup>ter</sup>.1 (e) PCT ergeht nicht.

Es wird nachdrücklich davon abgeraten, ein Sequenzprotokoll erst in diesem späten Stadium einzureichen. Zumindest bei umfangreichen Sequenzen wäre das EPA nicht in der Lage zu überprüfen, ob das Sequenzprotokoll über den Offenbarungsgehalt der internationalen Anmeldung im Anmeldezeitpunkt hinausgeht. Zu einer derartigen Überprüfung ist es nach Artikel 34 (2) b) und Regel 66.2 (a) iv) PCT zwar verpflichtet. Anders als bei der internationalen Recherche kann eine Erklärung des Anmelders, daß das nachgereichte Sequenzprotokoll nicht über den Offenbarungsgehalt der internationalen Anmeldung hinausgeht, das EPA nicht von einer eigenen Überprüfung dieses Tatbestandes (Artikel 34 (2) b), Regel 66.2 a) PCT) entbinden. Es gibt aber derzeit keine geeigneten Mittel, die einen zufriedenstellenden Vergleich der ursprünglichen Anmeldungsunterlagen mit dem nachgereichten Sequenzprotokoll zu vertretbarem Aufwand ermöglichen könnten. Daher könnte der internationale vorläufige Prüfungsbericht nicht auf der Grundlage des nachgereichten Sequenzprotokolls, sondern nur auf der Grundlage der am Anmeldetag in der Beschreibung vorhandenen Sequenzen erstellt werden.

#### IV. Eintritt der Euro-PCT-Anmeldungen in die regionale Phase vor dem EPA

1. Liegt ein nach Regel 5.2 PCT vorgeschriebenes Sequenzprotokoll bei Ablauf der internationalen Phase dem EPA als Bestimmungsamt oder ausgewähltem Amt nicht vor, entspricht es nicht dem Standard oder ist der vorgeschriebene Datenträger (siehe Standard, Nummern 39 ff.) nicht eingereicht worden, so wird der Anmelder aufgefordert, die festgestellten Mängel innerhalb einer **Frist von zwei Monaten** zu beheben (Regel 13<sup>ter</sup>.2 PCT, Regel 104b (3a) EPÜ und Artikel 5 des Beschlusses des Präsidenten). Gegebenenfalls wird der Anmelder aufgefordert, gleichzeitig zu erklären, daß das nachgereichte Sequenzprotokoll nicht über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht und/oder die auf dem

PCT). An invitation under Rule 13<sup>ter</sup>.1(e) PCT to file a sequence listing is not issued.

Applicants are strongly advised not to file sequence listings at this late stage. In the case of long sequence listings at least, the European Patent Office would not be in a position to examine whether the sequence listing goes beyond the content of the international application as disclosed at the time of filing, although under Article 34(2)(b) and Rule 66.2(a)(iv) PCT it is obliged to carry out such an examination. Unlike the situation pertaining to the international search, a statement from the applicant to the effect that the subsequently filed sequence listing does not go beyond the content of the international application as filed cannot release the EPO from its obligation to carry out its own examination (Article 34(2)(b), Rule 66.2(a) PCT). However, there is at present no appropriate means of carrying out a satisfactory comparison of the documents as originally filed with the subsequently filed sequence listing. The international preliminary examination report therefore could not be established on the basis of the subsequently filed sequence listing but only on the basis of the sequences available in the description on the date of filing.

#### IV. Entry of Euro-PCT applications into the regional phase before the EPO

1. If on expiry of the international phase the sequence listing prescribed in Rule 5.2 PCT is not available to the EPO as designated Office or elected Office or does not comply with the Standard, or if the prescribed data carrier (see Standard, paragraphs 39 ff) has not been filed, the applicant is invited to remedy such deficiencies within a **period of two months** (Rule 13<sup>ter</sup>.2 PCT, Rule 104b(3a) EPC and Article 5 of the decision of the President). Where appropriate, the applicant is invited at the same time to state that the subsequently filed sequence listing does not go beyond the content of the application as filed and/or that the information recorded on the data carrier is identical to the written sequence listing (Standard, para-

d'examen préliminaire international (règles 66.2 a) vii) et 70.12 iv) PCT). Il n'est pas envoyé d'invitation à déposer un listage des séquences conformément à la règle 13<sup>ter</sup>.1 e) PCT.

Il est tout à fait déconseillé de ne déposer un listage des séquences qu'à ce stade tardif de la procédure. Tout au moins dans le cas de séquences longues, l'OEB ne serait pas en mesure de vérifier si le listage des séquences ne s'étend pas au-delà de l'exposé de l'invention figurant dans la demande internationale à la date du dépôt. Il est certes tenu de procéder à une telle vérification en vertu de l'article 34 (2) b) et de la règle 66.2 a) iv) PCT, mais contrairement à ce qui se passe dans le cas de la recherche internationale, il ne suffit pas que le demandeur déclare que le listage des séquences déposé ultérieurement ne s'étend pas au-delà de l'exposé de l'invention figurant dans la demande internationale à la date du dépôt pour que l'OEB puisse se dispenser de procéder lui-même à cette vérification (article 34 (2) b), règle 66.2 a) PCT). Toutefois il n'existe pas actuellement de moyen approprié permettant d'établir une comparaison satisfaisante, ne nécessitant pas une somme de travail excessive, entre les pièces initiales de la demande et le listage des séquences déposé ultérieurement. C'est pourquoi le rapport d'examen préliminaire international ne saurait être établi sur la base du listage des séquences déposé ultérieurement, mais uniquement sur la base des séquences figurant dans la description à la date du dépôt.

#### IV. Entrée des demandes euro-PCT dans la phase régionale devant l'OEB

1. Si, au terme de la phase internationale, le listage des séquences visé à la règle 5.2 PCT n'est pas parvenu à l'OEB agissant en qualité d'Office désigné ou d'Office élu, n'a pas été établi conformément à la norme ou n'a pas été produit sur le support de données prescrit (cf. norme, paragraphes 39 s.), le demandeur est invité à remédier dans un **délai de deux mois** aux irrégularités constatées (règle 13<sup>ter</sup>.2 PCT, règle 104<sup>ter</sup> (3bis) CBE et article 5 de la décision du Président). Le cas échéant, le demandeur est invité à déclarer en même temps que le listage des séquences déposé ultérieurement ne s'étend pas au-delà du contenu de la demande telle qu'elle a été déposée et/ou que l'information figurant sur le support de données est identique à

Datenträger gespeicherte Information mit dem schriftlichen Sequenzprotokoll übereinstimmt (Standard, Nummern 37 und 40). Artikel 3 (2) und (3) des Beschlusses des Präsidenten gelten entsprechend (vgl. oben Punkt II 4.4).

2. Werden bestehende Mängel nach einer solchen Aufforderung nicht oder nicht rechtzeitig beseitigt, so wird die Anmeldung zurückgewiesen (Art. 91 (3) EPÜ).

3. Der Anmelder kann die Weiterbehandlung der Anmeldung nach Artikel 121 EPÜ beantragen.

4. Übersetzung gemäß Artikel 158 (2) EPÜ

Einer Übersetzung der Textstellen eines Sequenzprotokolls, das Bestandteil der Beschreibung ist, bedarf es grundsätzlich nicht, sofern das Sequenzprotokoll der Regel 12.1 (d) PCT und die Beschreibung der Regel 5.2 (b) PCT entspricht (Regel 49.5 (a-bis) PCT). Ist die Voraussetzung nach Regel 5.2 (b) PCT nicht erfüllt, so verlangt das EPA keine solche Übersetzung, wenn das Sequenzprotokoll in englisch abgefaßt ist und die internationale Anmeldung in einer anderen EPA-Amtssprache (deutsch oder französisch) veröffentlicht worden ist.

#### V. Veröffentlichung von Sequenzprotokollen durch das EPA – Akteneinsicht

1. Ist das Sequenzprotokoll **zusammen** mit den Unterlagen der europäischen Patentanmeldung (oder internationalen Anmeldung) als Teil der Beschreibung eingereicht worden, so wird es zusammen mit der Anmeldung durch das EPA (oder durch das Internationale Büro) veröffentlicht.

2. Wird ein Sequenzprotokoll zu einer europäischen oder zu einer in der regionalen Phase vor dem EPA anhängigen internationalen Anmeldung **nachgereicht**, so hängt seine Veröffentlichung von dem Zeitpunkt der Einreichung ab (Artikel 6 des Beschlusses des Präsidenten):

Wird es rechtzeitig vor Abschluß der technischen Vorbereitungen für die Veröffentlichung der europäischen Patentanmeldung<sup>5</sup> eingereicht, so wird es als Anlage zu den Anmeldeunterlagen (A-Schrift) oder im Falle einer Euro-PCT-Anmeldung, die nicht in einer Amtssprache des EPA

graphs 37 and 40). Article 3(2) and (3) of the decision of the President applies mutatis mutandis (see point II 4.4 above).

2. If existing deficiencies are not remedied or not remedied in due time after such an invitation, the application is refused (Art. 91(3) EPC).

3. The applicant may request further processing of the application under Article 121 EPC.

4. Translation in accordance with Article 158(2) EPC

There is no requirement to translate any text matter in a sequence listing which forms part of the description, provided that the sequence listing complies with Rule 12.1(d) PCT and the description with Rule 5.2(b) PCT (Rule 49.5(a-bis) PCT). If the requirement under Rule 5.2(b) PCT is not met, the EPO does not require such a translation where the sequence listing is in English and the international application has been published in one of the other official languages of the EPO (French or German).

#### V. Publication of sequence listings by the EPO – inspection of files

1. If the sequence listing has been filed **together** with the documents making up the European (or international) application as part of the description, it is published together with the application by the EPO (or the International Bureau).

2. If a sequence listing pertaining to a European patent application or to an international application in the regional phase before the EPO is **subsequently filed**, its publication depends on the date of filing (Article 6 of the decision of the President):

If it is filed in due time before completion of the technical preparations for publication of the European patent application<sup>5</sup>, it is published as an annex to the application documents (A document) or, in the case of a Euro-PCT application which has not been published internationally in one

celle que contient le listage imprimé (la norme, paragraphes 37 et 40). L'article 3(2) et (3) de la décision du Président est applicable par analogie (cf. supra, point II 4.4).

2. Si, après une telle invitation, le demandeur ne remédie pas aux irrégularités constatées ou n'y remédie pas en temps utile, la demande est rejetée (art. 91(3) CBE).

3. Le demandeur peut demander la poursuite de la procédure conformément à l'article 121 CBE.

4. Traduction conformément à l'article 158(2) CBE

Il n'est pas nécessaire en tout état de cause de déposer une traduction des éléments de texte d'un listage des séquences qui fait partie de la description, si ce listage des séquences est conforme à la règle 12.1 d) PCT et si la description est conforme à la règle 5.2 b) PCT (règle 49.5 (a-bis) PCT). Si la condition requise à la règle 5.2 b) PCT n'est pas remplie, l'OEB n'exige pas une telle traduction si le listage des séquences est rédigé en anglais et si la demande internationale a été publiée dans une autre langue officielle de l'OEB (allemand ou français).

#### V. Publication par l'OEB des listages des séquences – inspection publique

1. Si le listage des séquences a été déposé **en même temps** que les pièces de la demande européenne ou internationale comme partie de la description, il est publié avec la demande par l'OEB (ou par le Bureau international).

2. Si le listage des séquences d'une demande européenne ou d'une demande internationale se trouvant en instance devant l'OEB dans la phase régionale a été **déposé ultérieurement**, sa publication dépend de la date de son dépôt (article 6 de la décision du Président).

S'il est déposé en temps utile avant la fin des préparatifs techniques entrepris en vue de la publication de la demande de brevet européen<sup>5</sup>, il est publié en annexe aux pièces de la demande (document A) ou, s'il s'agit d'une demande euro-PCT qui n'a pas fait l'objet d'une publication interna-

<sup>5</sup> 7 Wochen vor Ablauf von 18 Monaten nach dem Anmelde- oder Prioritätstag; Beschluß des Präsidenten des EPA vom 14. Dezember 1992, ABI. EPA 1993, 55.

<sup>5</sup> Seven weeks before expiry of the eighteenth month following the date of filing or priority; Decision of the President of the EPO dated 14 December 1992, OJ EPO 1993, 55.

<sup>5</sup> Sept semaines avant l'expiration d'un délai de dix-huit mois à compter de la date de dépôt ou de priorité ; décision du Président de l'OEB en date du 14 décembre 1992, JO OEB 1993, 55.

international veröffentlicht worden ist, als Anlage zu der nach Artikel 158 (3) EPÜ veröffentlichten Übersetzung mitveröffentlicht.

In allen anderen Fällen besteht nur die Möglichkeit, bei Patenterteilung das Sequenzprotokoll zusammen mit der Patentschrift zu veröffentlichen. Die Veröffentlichung des nachgereichten Sequenzprotokolls zusammen mit dem gesonderten Recherchenbericht (sogenannte A3-Schrift) ist nicht vorgesehen.

3. Ein Sequenzprotokoll, das nicht zusammen mit den Anmeldeunterlagen oder mit der Patentschrift veröffentlicht wird, ist gegebenenfalls nur durch Akteneinsicht oder über die Patent-Sequenzdatenbanken der Öffentlichkeit zugänglich. Auf Antrag wird Akteneinsicht auch durch Herstellung und Zusendung einer Kopie des vom Anmelder eingereichten elektronischen Datenträgers gewährt (siehe II.2 und IV.1 und Regel 94 (3) EPÜ).

of the official languages of the EPO, as an annex to the translation published under Article 158(3) EPC.

In all other cases it is only possible to publish the sequence listing together with the patent specification when the patent is granted. There is no provision for publishing the subsequently filed sequence listing together with the separate search report (A3 document).

3. A sequence listing which is not published together with the application documents or with the patent specification is, where applicable, accessible to the public only by way of a file inspection or by way of the patent sequence databases. On request, inspection of files may also be granted by sending the requester a copy of the electronic data carrier filed by the applicant (see II.2 and IV.1 and Rule 94(3) EPC).

tionale dans une langue officielle de l'OEB, en annexe à la traduction publiée conformément à l'article 158(3) CBE.

Dans tous les autres cas, la seule possibilité consiste, si la demande donne lieu à la délivrance d'un brevet, à publier le listage des séquences avec le fascicule de brevet. Il n'est pas prévu qu'un listage des séquences déposé ultérieurement soit publié avec le rapport de recherche séparé (document A3).

3. Un listage des séquences qui n'est publié ni avec les pièces de la demande ni avec le fascicule de brevet n'est le cas échéant accessible au public que dans le cadre de l'inspection publique ou par l'intermédiaire de la banque de données relatives aux séquences contenues dans les brevets. Il est également procédé, sur requête, à l'inspection publique, par établissement et envoi d'une copie du support de données électronique produit par le demandeur (cf. II.2 et IV.1 et règle 94(3) CBE).

**Anlage zu der Mitteilung des Europäischen Patentamts vom 2. Oktober 1998 über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen\***

**Appendix to the Notice from the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings\***

**Annexe au Communiqué de l'Office européen des brevets en date du 2 octobre 1998, relatif à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences\***

**Freier Text des Sequenzprotokolls**

**Sequence Listing Free Text**

**Texte libre du listage des séquences**

**Beispiel 1:**

**Example 1:**

**Exemple 1 :**

Dies ist ein Beispiel des Abschnitts "Freier Text des Sequenzprotokolls" (vgl. Nummer 36 des WIPO-Standards ST. 25) in der Beschreibung einer in **Französisch** eingereichten europäischen oder internationalen Patentanmeldung, die ein Sequenzprotokoll mit Freitextelementen in Englisch zur Sequenz Nr. 3 enthält:

This is an example of a "Sequence Listing Free Text" section (see paragraph 36 of WIPO Standard ST. 25) in the description of a European patent application or international application filed in **French** and containing a sequence listing with English free text data elements for Sequence No. 3:

Ceci est un exemple de la rubrique particulière de la description intitulée "texte libre du listage des séquences" (voir le paragraphe 36 de la norme OMPI ST.25) pour une demande de brevet européen ou une demande internationale déposée en **français** et contenant un listage des séquences comportant des éléments de texte libre en anglais pour la séquence n° 3 :

```
<210> 3
<223> /note="PCV"
        /hôte spécifique="Sus scrofa"
        /type de tissu="rein"
        /lignée cellulaire="PK-15"
<223> /note="TAGTATTAC homologue au circovirus de plante et au motif nonanucléotide de géminivirus"
<223> /Codon de départ =1
        /note="similaire au circovirus BBTV de plante"
        /produit="site de réplication de la protéine associée au PCV"
<223> /note="fin de réplication de la protéine associée au PCV a lieu 6 nucléotides en aval du signal de polyadénylation"
<223> /note="TAGTATTAC homologue au circovirus de plante et au motif nonanucléotide de géminivirus"
```

**Beispiel 2:**

**Example 2**

**Exemple 2 :**

Dies ist ein Beispiel des Abschnitts "Freier Text des Sequenzprotokolls" (vgl. Nummer 36 des WIPO-Standards ST. 25) in der Beschreibung einer in **Deutsch** eingereichten europäischen oder internationalen Patentanmeldung, die ein Sequenzprotokoll mit Freitextelementen in Englisch zur Sequenz Nr. 3 enthält:

This is an example of a "Sequence Listing Free Text" section (see paragraph 36 of WIPO Standard ST. 25) in the description of a European patent application or international application filed in **German** and containing a sequence listing with English free text data elements for Sequence No. 3:

Ceci est un exemple de la rubrique particulière de la description intitulée "texte libre du listage des séquences" (voir le paragraphe 36 de la norme OMPI ST.25) pour une demande de brevet européen ou une demande internationale déposée en **allemand** et contenant un listage des séquences comportant des éléments de texte libre en anglais pour la séquence n° 3 :

```
<210> 3
<223> /Anmerkung="PCV"
        /Spezifischer Wirt="Sus scrofa"
        /Geweebetyp="Niere"
        /Zelllinie="PK-15"
<223> /Anmerkung="TAGTATTAC-Sequenz, homolog zu pflanzlichem Circovirus- und Geminivirus-spezifischem Nona-Nucleotidmotiv"
<223> /Startcodon = 1
        /Anmerkung="ähnlich dem pflanzlichen Circovirus BBTV"
        /Produkt="Replikationsstelle des assoziierten Proteins PCV"
<223> /Anmerkung="Ende des mutmaßlichen replikations-assoziierten Proteins PCV tritt 6 Nucleotide stromabwärts des potentiellen Polyadenylierungssignals auf"
<223> /Anmerkung="TAGTATTAC-Sequenz, homolog zu pflanzlichem Circovirus- und Geminivirus-spezifischem Nona-Nucleotidmotiv"
```

\* Siehe II 1.4 der Mitteilung in dieser Beilage, S. 60.

\* See II 1.4 of the Notice in this supplement, p. 60.

\* Voir le point II 1.4 du communiqué reproduit dans le présent supplément, p. 60.





## Beschluß des Präsidenten des Europäischen Patentamts vom 2. Oktober 1998 über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen

Der Präsident des Europäischen Patentamts, gestützt auf Artikel 10 EPÜ, Regeln 27a und 49 EPÜ sowie Regel 13<sup>ter</sup> PCT, beschließt:

### Artikel 1

*Standardisierte Darstellung von Nucleotid- und Aminosäuresequenzen in europäischen Patentanmeldungen*

Hat die in einer europäischen Patentanmeldung offenbarte Erfindung Nucleotid- oder Aminosäuresequenzen zum Gegenstand, so hat die Beschreibung ein Sequenzprotokoll zu enthalten, das dem im **Anhang** zu diesem Beschluß wiedergegebenen WIPO-Standard ST. 25 entspricht.

### Artikel 2

*Einreichung von Sequenzprotokollen auf elektronischem Datenträger*

(1) Zusätzlich ist ein Datenträger mit dem Sequenzprotokoll einzureichen, der dem im **Anhang** zu diesem Beschluß wiedergegebenen Standard entspricht.

(2) Wird ein Sequenzprotokoll nach seiner Einreichung beim Europäischen Patentamt berichtigt, so ist ein neuer Datenträger mit dem berichtigten Sequenzprotokoll einzureichen.

(3) Jedem Datenträger ist eine Erklärung des Anmelders beizufügen, daß die auf dem Datenträger gespeicherte Information mit dem schriftlichen Sequenzprotokoll übereinstimmt.

### Artikel 3

*Beseitigung von Mängeln*

(1) Wird den in Regel 27a EPÜ und in den Artikeln 1 und 2 dieses Beschlusses genannten Erfordernissen nicht entsprochen, so teilt die Eingangsstelle dies dem Anmelder mit und fordert ihn auf, die festgestellten Mängel innerhalb einer Frist von zwei Monaten zu beseitigen (Regel 41 (1) EPÜ).

## Decision of the President of the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings

The President of the European Patent Office, having regard to Article 10 EPC, Rules 27a and 49 EPC and Rule 13<sup>ter</sup> PCT, has decided as follows:

### Article 1

*Standardised presentation of nucleotide and amino acid sequences in European patent applications*

If the subject-matter of the invention disclosed in a European patent application is nucleotide or amino acid sequences, the description shall contain a sequence listing complying with WIPO Standard ST. 25 reproduced in the **annex** to this decision.

### Article 2

*Filing of sequence listings on electronic data carriers*

(1) In addition, a data carrier shall be filed containing the sequence listing and complying with the standard reproduced in the **annex** to this decision.

(2) If a sequence listing is corrected after being filed with the European Patent Office, a new data carrier containing the corrected sequence listing shall be filed.

(3) Each data carrier shall be accompanied by a statement by the applicant that the information recorded on the data carrier is identical to the written sequence listing.

### Article 3

*Rectification of deficiencies*

(1) If the requirements laid down in Rule 27a EPC and Articles 1 and 2 of this decision are not met, the Receiving Section shall inform the applicant accordingly and invite him to remedy the disclosed deficiencies within a period of two months (Rule 41(1) EPC).

## Décision du Président de l'Office européen des brevets, en date du 2 octobre 1998, relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences

Le Président de l'Office européen des brevets, vu l'article 10 CBE, les règles 27bis et 49 CBE ainsi que la règle 13<sup>ter</sup> PCT, décide :

### Article premier

*Présentation normalisée du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet européen*

Si l'invention exposée dans une demande de brevet européen a pour objet des séquences de nucléotides ou d'acides aminés, la description doit contenir un listage des séquences établi conformément à la norme ST. 25 de l'OMPI, reproduite dans l'**annexe** à la présente décision.

### Article 2

*Dépôt de listages des séquences sur support électronique de données*

(1) En outre, il y a lieu de déposer un support de données contenant un listage des séquences établi conformément à la norme reproduite dans l'**annexe** à la présente décision.

(2) Si un listage des séquences est rectifié après avoir été déposé auprès de l'Office européen des brevets, il y a lieu de déposer un nouveau support de données contenant le listage des séquences rectifié.

(3) Tout support de données doit être accompagné d'une déclaration du demandeur selon laquelle l'information figurant sur ce support est identique à celle que contient le listage écrit.

### Article 3

*Correction d'irrégularités*

(1) S'il n'est pas satisfait aux exigences visées à la règle 27bis CBE et aux articles premier et 2 de la présente décision, la section de dépôt le signale au demandeur et l'invite à remédier aux irrégularités constatées dans un délai de deux mois (règle 41(1) CBE).

(2) Die Frist wird einmal um höchstens zwei Monate verlängert, sofern der Anmelder dies vor Ablauf der Frist beantragt.

(3) Ist ein Datenträger beschädigt oder aus einem anderen Grund nicht lesbar, so teilt die Eingangsstelle dies dem Anmelder mit und fordert ihn auf, einen neuen, einwandfrei lesbaren Datenträger innerhalb einer von ihr zu bestimmenden Frist einzureichen.

#### Artikel 4

*Einreichung von Sequenzprotokollen beim EPA als Internationaler Recherchenbehörde*

In bezug auf internationale Anmeldungen nach dem PCT, für die nach Regel 5.2 PCT ein Sequenzprotokoll vorgeschrieben ist, verfährt das Europäische Patentamt als Internationale Recherchenbehörde nach Regel 13<sup>er</sup>.1 PCT und Abschnitten 208 und 513 mit Anhang C der Verwaltungsvorschriften zum PCT.

#### Artikel 5

*Einreichung von Sequenzprotokollen beim EPA als Bestimmungsamt oder als ausgewähltes Amt*

Bei Aufforderung zur Einreichung eines Sequenzprotokolls nach Regel 104b (3a) EPÜ ist Artikel 3 entsprechend anzuwenden.

#### Artikel 6

*Veröffentlichung von Sequenzprotokollen durch das EPA*

Ein am Anmeldetag der europäischen Patentanmeldung eingereichtes Sequenzprotokoll wird als Bestandteil der Beschreibung zusammen mit den Anmeldungsunterlagen veröffentlicht. Ein nach dem Anmeldetag eingereichtes oder berichtiges Sequenzprotokoll wird, wenn es vor dem Zeitpunkt vorliegt, an dem die technischen Vorbereitungen für die Veröffentlichung als abgeschlossen gelten, als Anlage zu den Anmeldungsunterlagen, andernfalls, bei Patenterteilung, als Anlage zur europäischen Patentschrift veröffentlicht.

(2) This period shall be extendable once only for a maximum of two months, provided the applicant requests such an extension prior to the expiry of the period.

(3) If a data carrier is damaged, or unreadable for some other reason, the Receiving Section shall inform the applicant accordingly and invite him to submit a flawless replacement within a period which it shall lay down.

#### Article 4

*Filing of sequence listings with the EPO as International Searching Authority*

With regard to international applications under the PCT, for which a sequence listing is prescribed under Rule 5.2 PCT, the European Patent Office shall act as International Searching Authority in accordance with Rule 13<sup>er</sup>.1 PCT and Sections 208 and 513 together with Annex C of the PCT Administrative Instructions.

#### Article 5

*Filing of sequence listings with the EPO as designated Office or elected Office*

Where an applicant is invited to file a sequence listing in accordance with Rule 104b(3a) EPC, Article 3 shall apply *mutatis mutandis*.

#### Article 6

*Publication of sequence listings by the EPO*

A sequence listing filed on the date of filing of the European patent application shall be published together with the application documents as part of the description. A sequence listing filed or corrected after the date of filing shall be published as an annex to the application documents, provided it is received before the date on which the technical preparations for publication are deemed to have been completed; otherwise, it shall be published as an annex to the European patent specification when the patent is granted.

(2) Le délai est prolongé une seule fois de deux mois au maximum, si le demandeur présente une requête en ce sens avant l'expiration du délai.

(3) Si un support de données est endommagé ou rendu illisible pour quelque autre raison, la section de dépôt le signale au demandeur et l'invite à déposer un nouveau support de données parfaitement lisible dans le délai qu'elle lui impartit.

#### Article 4

*Dépôt de listages des séquences auprès de l'OEB agissant en qualité d'administration chargée de la recherche internationale*

En ce qui concerne les demandes internationales déposées au titre du PCT pour lesquelles la règle 5.2 PCT prescrit un listage des séquences, l'Office européen des brevets, agissant en qualité d'administration chargée de la recherche internationale, procède conformément aux dispositions de la règle 13<sup>er</sup>.1 PCT et aux instructions administratives 208 et 513 du PCT, accompagnées de l'annexe C des Instructions administratives du PCT.

#### Article 5

*Dépôt de listages des séquences auprès de l'OEB agissant en tant qu'office désigné ou élu*

Lors de l'envoi d'une invitation à déposer un listage des séquences conformément à la règle 104ter (3bis) CBE, l'article 3 est applicable par analogie.

#### Article 6

*Publication par l'OEB de listages des séquences*

Un listage des séquences déposé à la date de dépôt de la demande de brevet européen est publié comme élément de la description avec les documents de la demande. Si un listage des séquences déposé ou rectifié après la date de dépôt est disponible avant la date à laquelle les préparatifs techniques entrepris en vue de la publication de la demande sont réputés achevés, il est publié en annexe aux documents de la demande, ou, si ce n'est pas le cas, lors de la délivrance du brevet, en annexe au fascicule du brevet européen.



**Artikel 7**

*Inkrafttreten*

Dieser Beschluß tritt am **1. Januar 1999** in Kraft und ersetzt den Beschluß vom 11. Dezember 1992, Beilage Nr. 2 zu ABI. EPA 12/1992, 1.

Geschehen zu München am  
2. Oktober 1998

*INGO KOBER*  
Präsident

**Article 7**

*Entry into force*

This decision shall enter into force on **1 January 1999** and shall replace the decision dated 11 December 1992, Supplement No. 2 to OJ EPO 12/1992, 1.

Done at Munich, 2 October 1998

*INGO KOBER*  
President

**Article 7**

*Entrée en vigueur*

La présente décision entre en vigueur le **1<sup>er</sup> janvier 1999** et remplace la décision du 11 décembre 1992, publiée dans le supplément n° 2 au Journal officiel 12/1992, 1.

Fait à Munich, le 2 octobre 1998

*INGO KOBER*  
Président

**ANHANG zum Beschluß des Präsidenten des Europäischen Patentamts vom 2. Oktober 1998 über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen**

**WIPO-Standard ST. 25**

**Standard für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in Patentanmeldungen**

Vom (ausführenden) PCIPI-Koordinierungsausschuß in seiner zweiundzwanzigsten Sitzung am 28. Mai 1998 angenommener Standard

Den Ämtern wird empfohlen, die Vorschriften des "Standards für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen nach dem Vertrag über die internationale Zusammenarbeit auf dem Gebiet des Patentwesens (PCT)", die in Anhang C zu den Verwaltungsvorschriften zum PCT enthalten sind, auch auf solche Patentanmeldungen anzuwenden, die keine internationalen Anmeldungen nach dem PCT sind, wobei zu beachten ist, daß bestimmte Vorschriften, die auf PCT-spezifische Verfahren und Erfordernisse zugeschnitten sind, auf Patentanmeldungen, die keine internationalen Anmeldungen nach dem PCT sind, möglicherweise keine Anwendung finden.\* Der Wortlaut des PCT-Standards ist nachfolgend wiedergegeben.

**ANNEX to the Decision of the President of the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings**

**WIPO Standard ST. 25**

**Standard for the presentation of nucleotide and amino acid sequence listings in patent applications**

Standard adopted by the PCIPI Executive Coordination Committee at its twenty-second session on May 28, 1998

It is recommended that Offices apply the provisions of the "Standard for the Presentation of Nucleotide and Amino Acid Sequence Listings in International Patent Applications Under the Patent Cooperation Treaty (PCT)" as set out in Annex C to the Administrative Instructions under the PCT, mutatis mutandis, to all patent applications other than the PCT international applications, noting that certain provisions specific to PCT procedures and requirements may not be applicable to patent applications other than PCT international applications.\* The text of that PCT Standard is reproduced on the following pages.

**ANNEXE à la décision du Président de l'Office européen des brevets, en date du 2 octobre 1998, relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences**

**Norme ST. 25 de l'OMPI**

**Norme relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet**

Norme adoptée par le Comité exécutif de coordination du PCIPI à sa vingt-deuxième session

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.\* Le texte de cette norme du PCT est reproduit ci-après.

\* Ist am 1. Juli 1998 die für ein Amt geltende nationale Gesetzgebung und Rechtspraxis nicht mit den Vorschriften der ersten beiden Sätze in Absatz 3 des "Standards für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen nach dem Vertrag über die internationale Zusammenarbeit auf dem Gebiet des Patentwesens (PCT)" vereinbar, so steht es diesem Amt frei, die betreffenden Vorschriften nicht zu befolgen, solange die Unvereinbarkeit besteht.

\* If, on July 1, 1998, the national law and practice applicable by an Office is not compatible with the provisions of the first two sentences of paragraph 3 of the "Standard for the Presentation of Nucleotide and Amino Acid Sequence Listings in International Patent Applications Under the Patent Cooperation Treaty (PCT)", that Office may choose not to follow those provisions for as long as that incompatibility continues.

\* Si, le 1<sup>er</sup> juillet 1998, la législation et la pratique nationales appliquées par un office ne sont pas compatibles avec les dispositions des deux premières phrases du paragraphe 3 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)", l'office peut choisir de ne pas appliquer ces dispositions tant que l'incompatibilité demeure.

## Anhang C zu den PCT-Verwaltungsvorschriften

Standard für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen nach dem PCT

### EINFÜHRUNG

1. Mit dem vorliegenden Standard soll auf eine standardisierte Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen hingewirkt werden. Anhand dieses Standards kann der Anmelder ein einheitliches Sequenzprotokoll erstellen, das von allen Anmeldeämtern, Internationalen Recherchenbehörden und mit der internationalen vorläufigen Prüfung beauftragten Behörden für die Zwecke der internationalen Phase und von allen Bestimmungsämtern und ausgewählten Ämtern für die Zwecke der nationalen Phase akzeptiert wird. Der Standard soll ferner eine genauere und qualitativ bessere Darstellung von Nucleotid- und Aminosäuresequenzen in internationalen Anmeldungen sicherstellen, die Darstellung und Verbreitung der Sequenzen zum Nutzen der Anmelder, der Öffentlichkeit und der Prüfer erleichtern, die Recherche von Sequenzdaten vereinfachen und den Austausch dieser Daten in elektronischer Form sowie ihre Aufnahme in computergestützte Datenbanken ermöglichen.

### DEFINITIONEN

2. Im Rahmen dieses Standards gelten folgende Definitionen:

i) Unter einem "Sequenzprotokoll" ist ein Teil der Beschreibung der Anmeldung in der eingereichten Fassung oder ein zur Anmeldung nachgereichtes Schriftstück zu verstehen, das die Nucleotid- und/oder Aminosäuresequenzen im einzelnen offenbart und sonstige verfügbare Angaben enthält.

ii) In das Protokoll dürfen nur unverzweigte Sequenzen von mindestens 4 Aminosäuren bzw. unverzweigte Sequenzen von mindestens 10 Nucleotiden aufgenommen werden. Verzweigte Sequenzen, Sequenzen mit weniger als 4 genau definierten Nucleotiden oder Aminosäuren sowie Sequenzen mit Nucleotiden oder Aminosäuren, die nicht in Anhang 2, Tabellen 1, 2, 3 und 4 aufgeführt sind, sind ausdrücklich ausgenommen.

iii) Unter "Nucleotiden" sind nur Nucleotide zu verstehen, die mittels der in Anhang 2, Tabelle 1 aufge-

## Annex C to the PCT Administrative Instructions

Standard for the presentation of nucleotide and amino acid sequence listings in international patent applications under the PCT

### INTRODUCTION

1. This Standard has been elaborated so as to provide standardization of the presentation of nucleotide and amino acid sequence listings in international patent applications. The Standard is intended to allow the applicant to draw up a single sequence listing which is acceptable to all receiving Offices, International Searching and Preliminary Examining Authorities for the purposes of the international phase, and to all designated and elected Offices for the purposes of the national phase. It is intended to enhance the accuracy and quality of presentations of nucleotide and amino acid sequences given in international applications, to make for easier presentation and dissemination of sequences for the benefit of applicants, the public and examiners, to facilitate searching of sequence data and to allow the exchange of sequence data in electronic form and the introduction of sequence data onto computerized databases.

### DEFINITIONS

2. For the purposes of this Standard:

(i) the expression "sequence listing" means a part of the description of the application as filed or a document filed subsequently to the application, which gives a detailed disclosure of the nucleotide and/or amino acid sequences and other available information;

(ii) sequences which are included are any unbranched sequences of four or more amino acids or unbranched sequences of ten or more nucleotides. Branched sequences, sequences with fewer than four specifically defined nucleotides or amino acids as well as sequences comprising nucleotides or amino acids other than those listed in Appendix 2, Tables 1, 2, 3 and 4, are specifically excluded from this definition;

(iii) "nucleotides" embrace only those nucleotides that can be represented using the symbols set forth

## Annexe C des Instructions administratives du PCT

Norme relative à la présentation du listage 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 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 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

fürten Symbole wiedergegeben werden können. Modifikationen wie z. B. methylierte Basen können nach der Anleitung in Anhang 2, Tabelle 2 beschrieben werden, sind aber in der Nucleotidsequenz nicht explizit auszuweisen.

iv) Unter "Aminosäuren" sind die in Anhang 2, Tabelle 3 aufgeführten gängigen L-Aminosäuren aus den natürlich vorkommenden Proteinen zu verstehen. Aminosäuresequenzen, die mindestens eine D-Aminosäure enthalten, fallen nicht unter diese Definition. Aminosäuresequenzen, die posttranslational modifizierte Aminosäuren enthalten, können mittels der in Anhang 2, Tabelle 3 aufgeführten Symbole wie die ursprünglich translatierte Aminosäure beschrieben werden, wobei die modifizierten Positionen wie z. B. Hydroxylierungen oder Glykosilierungen nach der Anleitung in Anhang 2, Tabelle 4 beschrieben werden, diese Modifikationen aber in der Aminosäuresequenz nicht explizit auszuweisen sind. Unter die vorstehende Definition fallen auch Peptide oder Proteine, die anhand der in Anhang 2, Tabelle 3 aufgeführten Symbole sowie einer an anderer Stelle aufgenommenen Beschreibung, die beispielsweise Aufschluß über ungewöhnliche Bindungen, Quervernetzungen (z. B. Disulfidbrücken) und "end caps", Nichtpeptidbindungen usw. gibt, als Sequenz wiedergegeben werden können.

v) Unter "Sequenzkennzahl" ist eine Zahl zu verstehen, die jeder im Protokoll aufgeführten Sequenz als SEQ ID NO zugewiesen wird.

vi) Die "numerische Kennzahl" ist eine dreistellige Zahl, die für ein bestimmtes Datenelement steht.

vii) Unter "sprachneutralem Vokabular" ist ein festes Vokabular zu verstehen, das im Sequenzprotokoll zur Wiedergabe der vom Hersteller einer Sequenzdatenbank vorgeschriebenen wissenschaftlichen Begriffe verwendet wird (dazu gehören wissenschaftliche Namen, nähere Bestimmungen und ihre Entsprechungen im festen Vokabular, die Symbole in Anhang 2, Tabellen 1, 2, 3 und 4 und die Merkmalschlüssel in Anhang 2, Tabellen 5 und 6).

viii) Unter "zuständiger Behörde" ist die Internationale Recherchenbehörde oder die mit der internationalen vorläufigen Prüfung beauftragte Behörde, die die internationale Recherche bzw. die internationale

in Appendix 2, Table 1. Modifications, for example, methylated bases, may be described as set forth in Appendix 2, Table 2, but shall not be shown explicitly in the nucleotide sequence;

(iv) "amino acids" are those L-amino acids commonly found in naturally occurring proteins and are listed in Appendix 2, Table 3. Those amino acid sequences containing at least one D-amino acid are not intended to be embraced by this definition. Any amino acid sequence that contains post-translationally modified amino acids may be described as the amino acid sequence that is initially translated using the symbols shown in Appendix 2, Table 3, with the modified positions, for example, hydroxylations or glycosylations, being described as set forth in Appendix 2, Table 4, but these modifications shall not be shown explicitly in the amino acid sequence. Any peptide or protein that can be expressed as a sequence using the symbols in Appendix 2, Table 3, in conjunction with a description elsewhere to describe, for example, abnormal linkages, cross-links (for example, disulfide bridge) and end caps, non-peptidyl bonds, etc., is embraced by this definition;

(v) "sequence identifier" is a unique integer that corresponds to the SEQ ID NO assigned to each sequence in the listing;

(vi) "numeric identifier" is a three-digit number which represents a specific data element;

(vii) "language-neutral vocabulary" is a controlled vocabulary used in the sequence listing that represents scientific terms as prescribed by sequence database providers (including scientific names, qualifiers and their controlled-vocabulary values, the symbols appearing in Appendix 2, Tables 1, 2, 3 and 4, and the feature keys appearing in Appendix 2, Tables 5 and 6);

(viii) "competent Authority" is the International Searching Authority that is to carry out the international search on the international application, or the International Preliminary Examining Authority that is to carry

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 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 de la recherche internationale ou l'administration chargée de l'examen préliminaire international pour la demande internationale

vorläufige Prüfung zu der internationalen Anmeldung durchführt, oder das Bestimmungsamt bzw. ausgewählte Amt zu verstehen, das mit der Bearbeitung der internationalen Anmeldung begonnen hat.

#### SEQUENZPROTOKOLL

3. Das Sequenzprotokoll im Sinne der Nummer 2 i) ist ans Ende der Anmeldung zu setzen, wenn es mit ihr zusammen eingereicht wird. Dieser Teil ist mit "Sequenzprotokoll" zu überschreiben, muß mit einer neuen Seite beginnen und sollte gesondert numeriert werden. Das Sequenzprotokoll ist Bestandteil der Beschreibung; vorbehaltlich Nummer 36 erübrigt es sich deshalb, die Sequenzen in der Beschreibung an anderer Stelle nochmals zu beschreiben.

4. Wird das Sequenzprotokoll im Sinne der Nummer 2 i) nicht zusammen mit der Anmeldung eingereicht, sondern als gesondertes Schriftstück nachgereicht (s. Nr. 37), so ist es mit der Überschrift "Sequenzprotokoll" und einer gesonderten Seitennummerierung zu versehen. Die in der Anmeldung in der eingereichten Fassung gewählte Numerierung der Sequenzen (s. Nr. 5) ist auch im nachgereichten Sequenzprotokoll beizubehalten.

5. Jeder Sequenz wird eine eigene Sequenzkennzahl zugeteilt. Die Kennzahlen beginnen mit 1 und setzen sich in aufsteigender Reihenfolge als ganze Zahlen fort. Gibt es zu einer Kennzahl keine Sequenz, so ist am Anfang der auf die SEQ ID NO folgenden Zeile unter der numerischen Kennzahl <400> der Code 000 anzugeben. Unter der numerischen Kennzahl <160> ist die Gesamtzahl der SEQ ID NOs anzugeben, und zwar unabhängig davon, ob im Anschluß an eine SEQ ID NO eine Sequenzkennzahl oder der Code 000 folgt.

6. In der Beschreibung, in den Ansprüchen und in den Zeichnungen der Anmeldung ist auf die im Sequenzprotokoll dargestellten Sequenzen mit "SEQ ID NO", gefolgt von der betreffenden Kennzahl, zu verweisen.

7. Für die Darstellung von Nucleotid- und Aminosäuresequenzen ist zumindest eine der folgenden drei Möglichkeiten zu wählen:

- i) nur Nucleotidsequenz
- ii) nur Aminosäuresequenz
- iii) Nucleotidsequenz zusammen mit der entsprechenden Aminosäuresequenz

out the international preliminary examination on the international application, or the designated/elected Office before which the processing of the international application has started.

#### SEQUENCE LISTING

3. The sequence listing as defined in paragraph 2(i) shall, where it is filed together with the application, be placed at the end of the application. This part shall be entitled "Sequence Listing," begin on a new page and preferably have independent page numbering. The sequence listing forms an integral part of the description; it is therefore unnecessary, subject to paragraph 36, to describe the sequences elsewhere in the description.

4. Where the sequence listing as defined in paragraph 2(i) is not contained in the application as filed but is a separate document furnished subsequently to the filing of the application (see paragraph 37), it shall be entitled "Sequence Listing" and shall have independent page numbering. The original numbering of the sequences (see paragraph 5) in the application as filed shall be maintained in the subsequently furnished sequence listing.

5. Each sequence shall be assigned a separate sequence identifier. The sequence identifiers shall begin with 1 and increase sequentially by integers. If no sequence is present for a sequence identifier, the code 000 should appear under numeric identifier <400>, beginning on the next line following the SEQ ID NO. The response for numeric identifier <160> shall include the total number of SEQ ID NOs, whether followed by a sequence or by the code 000.

6. In the description, claims or drawings of the application, the sequences represented in the sequence listing shall be referred to by the sequence identifier and preceded by "SEQ ID NO:".

7. Nucleotide and amino acid sequences should be represented by at least one of the following three possibilities:

- (i) a pure nucleotide sequence;
- (ii) a pure amino acid sequence;
- (iii) a nucleotide sequence together with its corresponding amino acid sequence.

en question, ou encore l'office désigné ou élu au sein duquel le traitement de la demande internationale a commencé.

#### LISTAGE DES SÉQUENCES

3. Le listage des séquences défini au paragraphe 2.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 "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 36, de décrire les séquences ailleurs dans la description.

4. Lorsque le listage des séquences défini au paragraphe 2.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 37), il doit s'intituler "Listage des séquences" et faire l'objet d'une pagination distincte. La numérotation des séquences (voir le paragraphe 5) figurant dans la demande telle qu'elle a été déposée doit être maintenue dans un listage des séquences remis ultérieurement.

5. A 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.

Bei einer Offenbarung im unter Ziffer iii genannten Format muß die Aminosäuresequenz als solche im Sequenzprotokoll mit einer eigenen Sequenzkennzahl gesondert offenbart werden.

For those sequences disclosed in the format specified in option (iii), above, the amino acid sequence must be disclosed separately in the sequence listing as a pure amino acid sequence with a separate integer sequence identifier.

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.

## NUCLEOTIDSEQUENZEN

### *Zu verwendende Symbole*

8. Nucleotidsequenzen sind nur anhand eines Einzelstrangs in Richtung vom 5'-Ende zum 3'-Ende von links nach rechts wiederzugeben. Die Begriffe 3' und 5' werden in der Sequenz nicht dargestellt.

9. Die Basen einer Nucleotidsequenz sind anhand der einbuchstabigen Codes für Nucleotidsequenzzeichen darzustellen. Es dürfen nur die in Anhang 2, Tabelle 1 aufgeführten Kleinbuchstaben verwendet werden.

10. Modifizierte Basen sind in der Sequenz selbst wie die entsprechenden unmodifizierten Basen oder mit "n" wiederzugeben, wenn die modifizierte Base zu den in Anhang 2, Tabelle 2 aufgeführten gehört; die Modifikation ist im Merkmalsteil des Sequenzprotokolls anhand der Codes in Anhang 2, Tabelle 2 näher zu beschreiben. Diese Codes dürfen in der Beschreibung oder im Merkmalsteil des Sequenzprotokolls, nicht jedoch in der Sequenz selbst verwendet werden (s. auch Nr. 32). Das Symbol "n" steht immer nur für ein einziges unbekanntes oder modifiziertes Nucleotid.

### *Zu verwendendes Format*

11. Bei Nucleotidsequenzen sind höchstens 60 Basen pro Zeile – mit einem Leerraum zwischen jeder Gruppe von 10 Basen – aufzuführen.

12. Die Basen einer Nucleotidsequenz (einschließlich Intronen) sind jeweils in Zehnergruppen aufzuführen; dies gilt nicht für die codierenden Teile der Sequenz. Bleiben am Ende nicht-codierender Teile einer Sequenz weniger als 10 Basen übrig, so sind sie zu einer Gruppe zusammenzufassen und durch einen Leerraum von angrenzenden Gruppen zu trennen.

13. Die Basen der codierenden Teile einer Nucleotidsequenz sind als Triplets (Codonen) aufzuführen.

## NUCLEOTIDE SEQUENCES

### *Symbols to Be Used*

8. A nucleotide sequence shall be presented only by a single strand, in the 5'-end to 3'-end direction from left to right. The terms 3' and 5' shall not be represented in the sequence.

9. The bases of a nucleotide sequence shall be represented using the one-letter code for nucleotide sequence characters. Only lower case letters in conformity with the list given in Appendix 2, Table 1, shall be used.

10. Modified bases shall be represented as the corresponding unmodified bases or as "n" in the sequence itself if the modified base is one of those listed in Appendix 2, Table 2, and the modification shall be further described in the feature section of the sequence listing, using the codes given in Appendix 2, Table 2. These codes may be used in the description or the feature section of the sequence listing but not in the sequence itself (see also paragraph 32). The symbol "n" is the equivalent of only one unknown or modified nucleotide.

### *Format to Be Used*

11. A nucleotide sequence shall be listed with a maximum of 60 bases per line, with a space between each group of 10 bases.

12. The bases of a nucleotide sequence (including introns) shall be listed in groups of 10 bases, except in the coding parts of the sequence. Leftover bases, fewer than 10 in number at the end of non-coding parts of a sequence, should be grouped together and separated from adjacent groups by a space.

13. The bases of the coding parts of a nucleotide sequence shall be listed as triplets (codons).

## 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. Die Zählung der Nucleotidbasen beginnt bei der ersten Base der Sequenz mit 1. Von hier aus ist die gesamte Sequenz in 5'-3'-Richtung fortlaufend durchzuzählen. Am rechten Rand ist jeweils neben der Zeile mit den einbuchstabigen Codes für die Basen die Nummer der letzten Base dieser Zeile anzugeben. Die vorstehend beschriebene Zählweise für Nucleotidsequenzen gilt auch für Nucleotidsequenzen mit ringförmiger Konfiguration, wobei allerdings die Bestimmung des ersten Nucleotids der Sequenz dem Anmelder überlassen bleibt.

15. Eine Nucleotidsequenz, die aus einem oder mehreren nichtbenachbarten Abschnitten einer größeren Sequenz oder aus Abschnitten verschiedener Sequenzen besteht, ist als gesonderte Sequenz mit eigener Sequenzkennzahl zu numerieren. Sequenzen mit einer oder mehreren Lücken sind als mehrere gesonderte Sequenzen mit eigenen Sequenzkennzahlen zu numerieren, wobei die Zahl der gesonderten Sequenzen der Zahl der jeweils zusammenhängenden Sequenzdatenreihen entspricht.

#### AMINOSÄURESEQUENZEN

##### *Zu verwendende Symbole*

16. Die Aminosäuren einer Protein- oder Peptidsequenz sind in Richtung von der Amino- zur Carboxylgruppe von links nach rechts aufzuführen, wobei die Amino- und Carboxylgruppen in der Sequenz nicht darzustellen sind.

17. Die Aminosäuren sind anhand des dreibuchstabigen Codes mit großem Anfangsbuchstaben entsprechend der Liste in Anhang 2, Tabelle 3 darzustellen. Eine Aminosäuresequenz, die einen Leerraum oder interne Terminatorsymbole (z. B. "Ter", "\*" oder ".") enthält, ist nicht als eine einzige Aminosäuresequenz, sondern als getrennte Aminosäuresequenzen darzustellen (s. Nr. 22).

18. Modifizierte und seltene Aminosäuren sind in der Sequenz selbst wie die entsprechenden unmodifizierten Aminosäuren oder mit "Xaa" wiederzugeben, wenn sie zu den in Anhang 2, Tabelle 4 aufgeführten gehören und die Modifikation im Merkmalsteil des Sequenzprotokolls anhand der Codes in Anhang 2, Tabelle 4 näher beschrieben wird. Diese Codes dürfen in der Beschreibung oder im Merkmalsteil des

14. The enumeration of the nucleotide shall start at the first base of the sequence with number 1. It shall be continuous through the whole sequence in the direction 5' to 3'. It shall be marked in the right margin, next to the line containing the one-letter codes for the bases, and giving the number of the last base of that line. The enumeration method for nucleotide sequences set forth above remains applicable to nucleotide sequences that are circular in configuration, with the exception that the designation of the first nucleotide of the sequence may be made at the option of the applicant.

15. A nucleotide sequence that is made up of one or more non-contiguous segments of a larger sequence or of segments from different sequences shall be numbered as a separate sequence, with a separate sequence identifier. A sequence with a gap or gaps shall be numbered as a plurality of separate sequences with separate sequence identifiers, with the number of separate sequences being equal in number to the number of continuous strings of sequence data.

#### AMINO ACID SEQUENCES

##### *Symbols to Be Used*

16. The amino acids in a protein or peptide sequence shall be listed in the amino to carboxy direction from left to right. The amino and carboxy groups shall not be represented in the sequence.

17. The amino acids shall be represented using the three-letter code with the first letter as a capital and shall conform to the list given in Appendix 2, Table 3. An amino acid sequence that contains a blank or internal terminator symbols (for example, "Ter" or "\*" or ".") may not be represented as a single amino acid sequence, but shall be presented as separate amino acid sequences (see paragraph 22).

18. Modified and unusual amino acids shall be represented as the corresponding unmodified amino acids or as "Xaa" in the sequence itself if the modified amino acid is one of those listed in Appendix 2, Table 4, and the modification shall be further described in the feature section of the sequence listing, using the codes given in Appendix 2, Table 4. These codes may be used in the description or the feature section of

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

Sequenzprotokolls, nicht jedoch in der Sequenz selbst verwendet werden (s. auch Nr. 32). Das Symbol "Xaa" steht immer nur für eine einzige unbekannte oder modifizierte Aminosäure.

the sequence listing but not in the sequence itself (see also paragraph 32). The symbol "Xaa" is the equivalent of only one unknown or modified amino acid.

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.

#### *Zu verwendendes Format*

#### *Format to Be Used*

#### *Mode de présentation à suivre*

19. Bei Protein- oder Peptidsequenzen sind höchstens 16 Aminosäuren pro Zeile mit einem Leerraum zwischen den einzelnen Aminosäuren aufzuführen.

19. A protein or peptide sequence shall be listed with a maximum of 16 amino acids per line, with a space provided between each amino acid.

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. Die den Codonen der codierenden Teile einer Nucleotidsequenz entsprechenden Aminosäuren sind unmittelbar unter den jeweiligen Codonen anzugeben. Wird ein Codon durch ein Intron aufgespalten, so ist das Aminosäuresymbol unter dem Teil des Codons anzugeben, der zwei Nucleotide enthält.

20. Amino acids corresponding to the codons in the coding parts of a nucleotide sequence shall be placed immediately under the corresponding codons. Where a codon is split by an intron, the amino acid symbol should be given below the portion of the codon containing two nucleotides.

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. Die Zählung der Aminosäuren beginnt bei der ersten Aminosäure der Sequenz mit 1. Fakultativ können die dem reifen Protein vorausgehenden Aminosäuren wie beispielsweise Präsequenzen, Prosequenzen, Präprosequenzen und Signalsequenzen, soweit vorhanden, mit negativem Vorzeichen numeriert werden, wobei die Rückwärtszählung mit der Aminosäure vor Nummer 1 beginnt. Null (0) wird nicht verwendet, wenn Aminosäuren zur Abgrenzung gegen das reife Protein mit negativem Vorzeichen numeriert werden. Die Nummern sind im Abstand von jeweils 5 Aminosäuren unter der Sequenz anzugeben. Die vorstehend beschriebene Zählweise für Aminosäuresequenzen gilt auch für Aminosäuresequenzen mit ringförmiger Konfiguration, wobei allerdings die Bestimmung der ersten Aminosäure der Sequenz dem Anmelder überlassen bleibt.

21. The enumeration of amino acids shall start at the first amino acid of the sequence, with number 1. Optionally, the amino acids preceding the mature protein, for example pre-sequences, pro-sequences, pre-pro-sequences and signal sequences, when present, may have negative numbers, counting backwards starting with the amino acid next to number 1. Zero (0) is not used when the numbering of amino acids uses negative numbers to distinguish the mature protein. It shall be marked under the sequence every five amino acids. The enumeration method for amino acid sequences set forth above remains applicable for amino acid sequences that are circular in configuration, with the exception that the designation of the first amino acid of the sequence may be made at the option of the applicant.

21. L'énumération des acides aminés doit commencer par le premier acide aminé de la séquence, qui portera le numéro 1. A titre facultatif, les acides aminés précédant la protéine mature, par exemple les pré-sequenzen, les pro-sequenzen et les pré-pro-sequenzen 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. Eine Aminosäuresequenz, die aus einem oder mehreren nichtbenachbarten Abschnitten einer größeren Sequenz oder aus Abschnitten verschiedener Sequenzen besteht, ist als gesonderte Sequenz mit eigener Sequenzkennzahl zu numerieren. Sequenzen mit einer oder mehreren Lücken sind als mehrere gesonderte Sequenzen mit eigenen Sequenzkennzahlen zu numerieren, wobei die Zahl der gesonderten Sequenzen der Zahl der jeweils zusammenhängenden Sequenzdatenreihen entspricht.

22. An amino acid sequence that is made up of one or more non-contiguous segments of a larger sequence or of segments from different sequences shall be numbered as a separate sequence, with a separate sequence identifier. A sequence with a gap or gaps shall be numbered as a plurality of separate sequences with separate sequence identifiers, with the number of separate sequences being equal in number to the number of continuous strings of sequence data.

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.



SONSTIGE VERFÜGBARE ANGABEN  
IM SEQUENZPROTOKOLL

23. Die Angaben sind im Sequenzprotokoll in der Reihenfolge anzugeben, in der sie in der Liste der numerischen Kennzahlen der Datenelemente in Anhang 1 aufgeführt sind.

24. Für die Angaben im Sequenzprotokoll sind nur die in Anhang 1 aufgeführten numerischen Kennzahlen, nicht aber die dazugehörigen Beschreibungen zu verwenden. Die Angaben müssen unmittelbar auf die numerische Kennzahl folgen; im Sequenzprotokoll brauchen nur diejenigen numerischen Kennzahlen angegeben zu werden, zu denen auch Angaben vorliegen. Die einzigen beiden Ausnahmen zu dieser Vorschrift bilden die numerischen Kennzahlen <220> und <300>, die für die Rubrik "Merkmal" bzw. "Veröffentlichungsangaben" stehen und mit den Angaben unter den numerischen Kennzahlen <221> bis <223> bzw. <301> bis <313> zusammenhängen. Werden unter diesen numerischen Kennzahlen im Sequenzprotokoll Angaben zu den Merkmalen oder zur Veröffentlichung gemacht, so sollte auch die numerische Kennzahl <220> bzw. <300> aufgeführt, die dazugehörige Rubrik aber nicht ausgefüllt werden. Generell sollte zwischen den numerischen Kennzahlen eine Leerzeile eingefügt werden, wenn sich die an erster oder zweiter Position der numerischen Kennzahl stehende Ziffer ändert. Eine Ausnahme von dieser allgemeinen Regel bildet die numerische Kennzahl <310>, der keine Leerzeile vorausgehen darf. Auch vor jeder Wiederholung der numerischen Kennzahl ist eine Leerzeile einzufügen.

*Obligatorische Datenelemente*

25. In das Sequenzprotokoll sind außerdem vor der eigentlichen Nucleotid- und/oder Aminosäuresequenz die folgenden in Anhang 1 definierten Angaben (obligatorische Datenelemente) aufzunehmen:

<110> Name des Anmelders  
<120> Bezeichnung der Erfindung  
<160> Anzahl der SEQ ID NOs  
<210> SEQ ID NO: x  
<211> Länge  
<212> Art  
<213> Organismus  
<400> Sequenz

OTHER AVAILABLE INFORMATION IN  
THE SEQUENCE LISTING

23. The order of the items of information in the sequence listings shall follow the order in which those items are listed in the list of numeric identifiers of data elements as defined in Appendix 1.

24. Only numeric identifiers of data elements as defined in Appendix 1 shall be used for the presentation of the items of information in the sequence listing. The corresponding numeric identifier descriptions shall not be used. The provided information shall follow immediately after the numeric identifier while only those numeric identifiers for which information is given need appear on the sequence listing. Two exceptions to this requirement are numeric identifiers <220> and <300>, which serve as headers for "Feature" and "Publication Information," respectively, and are associated with information in numeric identifiers <221> to <223> and <301> to <313>, respectively. When feature and publication information is provided in the sequence listing under those numeric identifiers, numeric identifiers <220> and <300>, respectively, should be included, but left blank. Generally, a blank line shall be inserted between numeric identifiers when the digit in the first or second position of the numeric identifier changes. An exception to this general rule is that no blank line should appear preceding numeric identifier <310>. Additionally, a blank line shall precede any repeated numeric identifier.

*Mandatory Data Elements*

25. The sequence listing shall include, in addition to and immediately preceding the actual nucleotide and/or amino acid sequence, the following items of information defined in Appendix 1 (mandatory data elements):

<110> Applicant name  
<120> Title of invention  
<160> Number of SEQ ID NOs  
<210> SEQ ID NO: x  
<211> Length  
<212> Type  
<213> Organism  
<400> Sequence

AUTRES RENSEIGNEMENTS  
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.

*Eléments 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

Ist der Name des Anmelders (numerische Kennzahl <110>) nicht in Buchstaben des lateinischen Alphabets geschrieben, so ist er – im Wege der Transliteration oder der Übersetzung ins Englische – auch in lateinischen Buchstaben anzugeben.

Where the name of the applicant (numeric identifier <110>) is written in characters other than those of the Latin alphabet, it shall also be indicated in characters of the Latin alphabet either as a mere transliteration or through translation into English.

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.

Die Datenelemente mit Ausnahme der Angaben unter den numerischen Kennzahlen <110>, <120> und <160> sind für jede im Sequenzprotokoll aufgeführte Sequenz zu wiederholen. Gibt es zu einer Sequenzkennzahl keine Sequenz, so müssen nur die Datenelemente unter den numerischen Kennzahlen <210> und <400> angegeben werden (s. Nr. 5 und SEQ ID NO: 4 in dem in Anhang 3 dieses Standards enthaltenen Beispiel).

The data elements, except those under numeric identifiers <110>, <120> and <160>, shall be repeated for each sequence included in the sequence listing. Only the data elements under numeric identifiers <210> and <400> are mandatory if no sequence is present for a sequence identifier (see paragraph 5, above, and SEQ ID NO: 4 in the example depicted in Appendix 3 of this Standard).

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. In Sequenzprotokolle, die zusammen mit der dazugehörigen Anmeldung oder vor Vergabe einer Anmeldeungsnummer eingereicht werden, ist neben den unter Nummer 25 genannten Datenelementen auch das folgende aufzunehmen:

26. In addition to the data elements identified in paragraph 25, above, when a sequence listing is filed at the same time as the application to which it pertains or at any time prior to the assignment of an application number, the following data element shall be included in the sequence listing:

26. Outre les éléments de données indiqués au paragraphe 25, 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> Aktenzeichen

<130> File reference

<130> Référence du dossier

27. In Sequenzprotokolle, die auf Anforderung einer zuständigen Behörde oder nach Vergabe einer Anmeldeungsnummer eingereicht werden, sind neben den unter Nummer 25 genannten Datenelementen auch die folgenden aufzunehmen:

27. In addition to the data elements identified in paragraph 25, above, when a sequence listing is filed in response to a request from a competent Authority or at any time following the assignment of an application number, the following data elements shall be included in the sequence listing:

27. Outre les éléments de données énumérés au paragraphe 25, 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> Vorliegende Patentanmeldung  
<141> Anmeldetag der vorliegenden Anmeldung

<140> Current patent application  
<141> Current filing date

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

28. In Sequenzprotokolle, die zu einer Anmeldung eingereicht werden, die die Priorität einer früheren Anmeldung in Anspruch nimmt, sind neben den unter Nummer 25 genannten Datenelementen auch die folgenden aufzunehmen:

28. In addition to the data elements identified in paragraph 25, above, when a sequence listing is filed relating to an application which claims the priority of an earlier application, the following data elements shall be included in the sequence listing:

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> Frühere Patentanmeldung

<150> Earlier patent application

<150> Demande de brevet antérieure

<151> Anmeldetag der früheren Anmeldung

<151> Earlier application filing date

<151> Date de dépôt de la demande antérieure

29. Wird in der Sequenz "n", "Xaa" oder eine modifizierte Base oder modifizierte/seltene L-Aminosäure aufgeführt, so müssen die folgenden Datenelemente angegeben werden:

<220> Merkmal  
<221> Name/Schlüssel  
<222> Lage  
<223> Sonstige Informationen

30. Ist der Organismus (numerische Kennzahl <213>) eine "künstliche Sequenz" oder "unbekannt", so müssen die folgenden Datenelemente angegeben werden:

<220> Merkmal  
<223> Sonstige Angaben

#### *Fakultative Datenelemente*

31. Alle in Anhang 1 definierten Datenelemente, die unter den Nummern 25 bis 30 nicht erwähnt sind, sind fakultativ (fakultative Datenelemente).

#### *Angabe von Merkmalen*

32. Merkmale, die (unter der numerischen Kennzahl <220>) zu einer Sequenz angegeben werden, sind durch die in Anhang 2, Tabellen 5 und 6<sup>1</sup> aufgeführten "Merkmal-schlüssel" zu beschreiben.

#### *Freier Text*

33. Unter "freiem Text" ist eine verbale Beschreibung der Eigenschaften der Sequenz ohne Verwendung des sprachneutralen Vokabulars im Sinne der Nummer 2 (vii) unter der numerischen Kennzahl <223> (Sonstige Angaben) zu verstehen.

34. Der freie Text sollte sich auf einige kurze, für das Verständnis der Sequenz unbedingt notwendige Begriffe beschränken. Er sollte in englischer Sprache für jedes Datenelement nicht länger als 4 Zeilen sein und höchstens 65 Buchstaben pro Zeile umfassen. Alle weiteren Angaben sind in den Hauptteil der Beschreibung in der dort verwendeten Sprache aufzunehmen.

35. Freier Text sollte in englischer Sprache abgefaßt sein.

36. Enthält das Sequenzprotokoll, das Bestandteil der Beschreibung ist, freien Text, so ist dieser im Hauptteil der Beschreibung in der dort verwen-

29. If "n" or "Xaa" or a modified base or modified/unusual L-amino acid is used in the sequence, the following data elements are mandatory:

<220> Feature  
<221> Name/key  
<222> Location  
<223> Other information

30. If the organism (numeric identifier <213>) is "Artificial Sequence" or "Unknown," the following data elements are mandatory:

<220> Feature  
<223> Other information

#### *Optional Data Elements*

31. All data elements defined in Appendix 1, not mentioned in paragraphs 25 to 30, above, are optional (optional data elements).

#### *Presentation of Features*

32. When features of sequences are presented (that is, numeric identifier <220>), they shall be described by the "feature keys" set out in Appendix 2, Tables 5 and 6.<sup>1</sup>

#### *Free Text*

33. "Free text" is a wording describing characteristics of the sequence under numeric identifier <223> (Other information) which does not use language-neutral vocabulary as referred to in paragraph 2(vii).

34. The use of free text shall be limited to a few short terms indispensable for the understanding of the sequence. It shall not exceed four lines with a maximum of 65 characters per line for each given data element, when written in English. Any further information shall be included in the main part of the description in the language thereof.

35. Any free text should preferably be in the English language.

36. Where the sequence listing part of the description contains free text, any such free text shall be repeated in the main part of the description in

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

#### *Eléments 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.

#### *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.

36. Lorsque, dans la description, la partie réservée au listage des séquences contient du texte libre, celui-ci doit être répété dans la partie

<sup>1</sup> Diese Tabellen enthalten Auszüge aus den Merkmals-tabellen "DDBJ/EMBL/GenBank Feature Table" (Nucleotidsequenzen) und "SWISS PROT Feature Table" (Aminosäuresequenzen).

<sup>1</sup> These tables contain extracts of the DDBJ/EMBL/GenBank Feature Table (nucleotide sequences) and the SWISS PROT Feature Table (amino acid sequences).

<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).

deten Sprache zu wiederholen. Es wird empfohlen, den in der Sprache des Hauptteils der Beschreibung abgefaßten freien Text in einen besonderen Abschnitt der Beschreibung mit der Überschrift "Sequenzprotokoll – freier Text" aufzunehmen.

#### NACHGERECHTES SEQUENZPROTOKOLL

37. Alle Sequenzprotokolle, die nicht zusammen mit der Anmeldung, sondern später eingereicht werden, dürfen nicht über den Offenbarungsgehalt der Anmeldung in der eingereichten Fassung hinausgehen; ihnen ist eine entsprechende Erklärung beizufügen. Die zur Anmeldung nachgereichten Sequenzprotokolle dürfen also nur die Sequenzen enthalten, die in der Anmeldung in der eingereichten Fassung offenbart worden sind.

38. Sequenzprotokolle, die nicht zusammen mit der Anmeldung eingereicht worden sind, sind nicht Bestandteil der Anmeldung. Da die Regeln 13<sup>ter</sup>, 26.3 und 91 sowie der Artikel 34 PCT Anwendung finden, besteht jedoch – vorbehaltlich der einschlägigen Bestimmungen – die Möglichkeit, Sequenzprotokolle, die zusammen mit der Anmeldung eingereicht wurden, im Wege der Mängelbeseitigung gemäß Regel 13<sup>ter</sup> oder 26.3 PCT und – bei Vorliegen offensichtlicher Fehler – gemäß Regel 91 PCT zu berichtigen oder gemäß Artikel 34 PCT zu ändern sowie ferner im Wege der Änderung der Anmeldung gemäß Artikel 34 PCT ein Sequenzprotokoll nachzureichen.

#### SEQUENZPROTOKOLL IN MASCHINENLESBARER FORM

39. Das in der Anmeldung enthaltene Sequenzprotokoll ist zusätzlich auch in maschinenlesbarer Form einzureichen, wenn die zuständige Behörde dies verlangt.

40. Das zusätzlich eingereichte maschinenlesbare Sequenzprotokoll muß mit dem geschriebenen Protokoll identisch sein und ist zusammen mit einer Erklärung folgenden Wortlauts einzureichen: "Die in maschinenlesbarer Form aufgezeichneten Angaben sind mit dem geschriebenen Sequenzprotokoll identisch."

the language thereof. It is recommended that the free text in the language of the main part of the description be put in a specific section of the description called "Sequence Listing Free Text."

#### SUBSEQUENTLY FURNISHED SEQUENCE LISTING

37. Any sequence listing which is not contained in the application as filed but which is furnished subsequently shall not go beyond the disclosure in the application as filed and shall be accompanied by a statement to that effect. This means that a sequence listing furnished subsequently to the filing of the application shall contain only those sequences that were disclosed in the application as filed.

38. Any sequence listing not contained in the application as filed does not form part of the application. However, the provisions of PCT Rules 13<sup>ter</sup>, 26.3 and 91 and PCT Article 34 would apply, so that it may be possible, subject to the applicable provisions, for a sequence listing contained in the application as filed to be corrected under PCT Rules 13<sup>ter</sup> or 26.3, rectified under PCT Rule 91 (in the case of an obvious error), or amended under PCT Article 34, or for a sequence listing to be submitted under PCT Article 34 as an amendment to the application.

#### COMPUTER READABLE FORM OF THE SEQUENCE LISTING

39. A copy of the sequence listing shall also be submitted in computer readable form, in addition to the sequence listing as contained in the application, whenever this is required by the competent Authority.

40. Any sequence listing in computer readable form submitted in addition to the written sequence listing shall be identical to the written sequence listing and shall be accompanied by a statement that "the information recorded in computer readable form is identical to the written sequence listing."

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 "texte libre du listage des séquences".

#### LISTAGE DES SÉQUENCES DÉPOSÉ ULTÉRIEUREMENT

37. 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.

38. Tout listage des séquences ne figurant pas dans la demande telle qu'elle a été déposée ne fait pas partie de la demande. Toutefois, les dispositions des règles 13<sup>er</sup>, 26.3 et 91 du règlement d'exécution du PCT et celles de l'article 34 du PCT s'appliquent normalement, de telle sorte qu'il est possible, sous réserve des dispositions applicables, qu'un listage des séquences figurant dans la demande telle qu'elle a été déposée fasse l'objet d'une correction en vertu de la règle 13<sup>er</sup> ou 26.3 du règlement d'exécution du PCT, d'une rectification en vertu de la règle 91 du règlement d'exécution du PCT (dans le cas d'une erreur évidente) ou d'une modification en vertu de l'article 34 du PCT, ou qu'un listage des séquences soit soumis en vertu de l'article 34 du PCT à titre de modification de la demande.

#### LISTAGE DES SÉQUENCES SOUS UNE FORME DÉCHIFFRABLE PAR ORDINATEUR

39. Outre le listage des séquences figurant dans la demande, une copie de ce même listage doit être fournie sous une forme déchiffable par ordinateur chaque fois que l'administration compétente l'exige.

40. Tout listage des séquences sous une forme déchiffable 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 déchiffable par ordinateur sont identiques à celles du listage des séquences présenté par écrit."

41. Das gesamte ausdrückbare Exemplar des Sequenzprotokolls muß in einer einzigen Datei enthalten sein, die nach Möglichkeit auf einer einzigen Diskette oder einem sonstigen von der zuständigen Behörde akzeptierten elektronischen Datenträger aufgezeichnet sein soll. Diese Datei ist mittels der IBM<sup>2</sup>-Codetabellen (Code Page) 437, 932<sup>3</sup> oder einer kompatiblen Codetabelle zu codieren. Eine Codetabelle, wie sie z. B. für japanische, chinesische, kyrillische, arabische, griechische oder hebräische Schriftzeichen benötigt wird, gilt als kompatibel, wenn sie das lateinische Alphabet und die arabischen Ziffern denselben Hexadezimalpositionen zuordnet wie die genannten Codetabellen.

42. Die maschinenlesbare Fassung soll mit spezieller Software wie PatentIn oder anderen kundenspezifischen Computerprogrammen erstellt werden; sie kann mit beliebigen Mitteln erstellt werden, solange das Sequenzprotokoll auf einer Diskette oder einem anderen von der zuständigen Behörde akzeptierten elektronischen Träger eingereicht wird und mit einem Betriebssystem für Personal Computer gelesen werden kann, das von der zuständigen Behörde akzeptiert wird.

43. Bei Verwendung von Disketten ist eine Datenkompression erlaubt, sofern die komprimierte Datei in einem selbstextrahierenden Format erstellt worden ist, das sich auf einem von der zuständigen Behörde akzeptierten Betriebssystem für Personal Computer selbst dekomprimiert.

44. Auf der Diskette oder einem anderen von der zuständigen Behörde akzeptierten elektronischen Träger ist ein Etikett fest anzubringen, auf dem von Hand oder mit der Maschine in Blockschrift der Name des Anmelders, die Bezeichnung der Erfindung, ein Aktenzeichen, der Zeitpunkt der Aufzeichnung der Daten, das Computerbetriebssystem und der Name der zuständigen Behörde eingetragen sind.

45. Wird die Diskette oder ein anderes von der zuständigen Behörde akzeptiertes elektronisches Medium erst nach dem Tag der Anmeldung

41. The entire printable copy of the sequence listing shall be contained within one electronic file preferably on a single diskette or any other electronic medium that is acceptable to the competent Authority. The file recorded on the diskette or any other electronic medium that is acceptable to the competent Authority shall be encoded using IBM<sup>2</sup> Code Page 437, IBM Code Page 932 or a compatible code page. A compatible code page, as would be required for, for example, Japanese, Chinese, Cyrillic, Arabic, Greek or Hebrew characters, is one that assigns the Roman alphabet and numerals to the same hexadecimal positions as do the specified code pages.

42. The computer readable form shall preferably be created by dedicated software such as PatentIn or other custom computer programs; it may be created by any means, as long as the sequence listing on a submitted diskette or any other electronic medium that is acceptable to the competent Authority is readable under a Personal Computer Operating system that is acceptable to the competent Authority.

43. File compression is acceptable when using diskette media, so long as the compressed file is in a self-extracting format that will decompress on a Personal Computer Operating system that is acceptable to the competent Authority.

44. The diskette or any other electronic medium that is acceptable to the competent Authority shall have a label permanently affixed thereto on which has been hand-printed, in block capitals or typed, the name of the applicant, the title of the invention, a reference number, the date on which the data were recorded, the computer operating system and the name of the competent Authority.

45. If the diskette or any other electronic medium that is acceptable to the competent Authority is submitted after the date of filing of an applica-

41. 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'administration compétente. 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 ou une page de code compatible. 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.

42. Le listage déchiffirable par ordinateur doit, de préférence, être créé par un logiciel spécialisé tel que PatentIn ou d'autres programmes informatiques personnalisés; il peut être créé par tout autre moyen dans la mesure où le listage des séquences figurant sur la disquette ou sur tout autre support électronique admis par l'administration compétente est déchiffirable sous un système d'exploitation d'ordinateur individuel lui aussi admis par l'administration compétente.

43. La compression d'un fichier est admise lorsque le support est une disquette, dans la mesure où le fichier compressé se présente sous un format auto-extractible qui se décompressera sur un système d'exploitation d'ordinateur personnel admis par l'administration compétente.

44. Sur la disquette ou sur tout autre support électronique admis par l'administration compétente doit être apposée 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, le système d'exploitation informatique et le nom de l'administration compétente.

45. Si la disquette ou tout autre support électronique admis par l'administration compétente est fournie après la date de dépôt de la

<sup>2</sup> IBM ist eine für die International Business Machine Corporation, Vereinigte Staaten von Amerika, eingetragene Handelsmarke.

<sup>3</sup> Die genannten Codetabellen gelten für Personal Computer praktisch als Standard.

<sup>2</sup> IBM is a registered trademark of International Business Machine Corporation, United States of America.

<sup>3</sup> The specified code pages are *de facto* standards for personal computers.

<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.

eingereicht, so sind auf den Etiketten auch der Anmeldetag und die Anmelde­nummer anzugeben.

46. Jeder Berichtigung des geschriebenen Sequenzprotokolls nach Regel 13<sup>ter</sup>.1 (a) i) oder 26.3 PCT, jeder Berichtigung eines offensichtlichen Fehlers im geschriebenen Sequenzprotokoll nach Regel 91 PCT und jeder nach Artikel 34 PCT vorgenommenen Änderung, die sich auf ein geschriebenes Sequenzprotokoll bezieht, ist ein maschinenlesbares Sequenzprotokoll beizufügen, in dem diese Berichtigung oder Änderung berücksichtigt worden ist.

tion, the labels shall also include the filing date of the application and the application number.

46. Any correction of the written sequence listing which is submitted under PCT Rules 13<sup>ter</sup>.1(a)(i) or 26.3, any rectification of an obvious error in the written sequence listing which is submitted under PCT Rule 91, or any amendment which includes a written sequence listing and which is submitted under PCT Article 34, shall be accompanied by a computer readable form of the sequence listing including any such correction, rectification or amendment.

demande, cette date et le numéro de la demande doivent aussi figurer sur les étiquettes.

46. Toute correction – en vertu des règles 13<sup>er</sup>.1.a)i) ou 26.3 du règlement d'exécution du PCT – du listage des séquences présenté par écrit, toute rectification – en vertu de la règle 91 du règlement d'exécution du PCT – d'une erreur évidente dans le listage des séquences présenté par écrit ou toute modification – en vertu de l'article 34 du PCT – qui comprend un listage des séquences présenté par écrit doit être accompagnée du listage des séquences sous une forme déchiffrable par ordinateur, comprenant la correction, la rectification ou la modification en question.

**Anhang 1****Numerische Kennzahlen**

In Sequenzprotokollen, die zu Anmeldungen eingereicht werden, dürfen nur die nachstehenden numerischen Kennzahlen verwendet werden. Die Überschriften der nachstehenden Datenelementrubriken dürfen in den Sequenzprotokollen nicht erscheinen.

Die numerischen Kennzahlen der obligatorischen Datenelemente, d. h. der Datenelemente, die in alle Sequenzprotokolle aufgenommen werden müssen (s. Nr. 25 dieses Standards: Kennziffern 110, 120, 160, 210, 211, 212, 213 und 400), und die numerischen Kennzahlen der Datenelemente, die in den in diesem Standard genannten Fällen aufgenommen werden müssen (s. Nr. 26,

27, 28, 29 und 30 dieses Standards: Kennzahlen 130, 140, 141, 150 und 151 sowie 220 bis 223), sind mit dem Buchstaben "O" gekennzeichnet.

Die numerischen Kennzahlen der fakultativen Datenelemente (s. Nr. 31 dieses Standards) sind mit dem Buchstaben "F" gekennzeichnet.

Numerische Kennzahl	Numerische Kennzahl Beschreibung	Obligatorisch (O) oder fakultativ (F)	Bemerkungen
<110>	Name des Anmelders	O	Ist der Name des Anmelders nicht in lateinischen Buchstaben geschrieben, so muß er – im Wege der Transliteration oder der Übersetzung ins Englische – auch in lateinischen Buchstaben angegeben werden.
<120>	Bezeichnung der Erfindung	O	
<130>	Aktenzeichen	O in den Fällen nach Nr. 26 Standard	Siehe Nr. 26 Standard.
<140>	Vorliegende Patentanmeldung	O in den Fällen nach Nr. 27 Standard	Siehe Nr. 27 Standard; die vorliegende Patentanmeldung ist zu kennzeichnen durch den zweibuchstabigen Code nach dem WIPO-Standard ST. 3, gefolgt von der Anmeldungsnummer (in dem Format, das von der Behörde für gewerblichen Rechtsschutz verwendet wird, bei der diese Patentanmeldung eingereicht wird) oder – bei internationalen Anmeldungen – von der internationalen Anmeldungsnummer.
<141>	Anmeldetag der vorliegenden Anmeldung	O in den Fällen nach Nr. 27 Standard	Siehe Nr. 27 Standard; das Datum ist entsprechend dem WIPO-Standard ST. 2 anzugeben (CCYY MM DD).
<150>	Frühere Patentanmeldung	O in den Fällen nach Nr. 28 Standard	Siehe Nr. 28 Standard; die frühere Patentanmeldung ist zu kennzeichnen durch den zweibuchstabigen Code entsprechend dem WIPO-Standard ST. 3, gefolgt von der Anmeldungsnummer (in dem Format, das von der Behörde für gewerblichen Rechtsschutz verwendet wird, bei der die frühere Patentanmeldung eingereicht wurde) oder – bei internationalen Anmeldungen – von der internationalen Anmeldungsnummer.
<151>	Anmeldetag der früheren Anmeldung	O in den Fällen nach Nr. 28 Standard	Siehe Nr. 28 Standard; das Datum ist entsprechend dem WIPO-Standard ST. 2 anzugeben (CCYY MM DD).
<160>	Anzahl der SEQ ID NOs	O	
<170>	Software	F	
<210>	Angaben zu SEQ ID NO: x	O	Anzugeben ist eine ganze Zahl, die die SEQ ID NO darstellt.
<211>	Länge	O	Sequenzlänge, ausgedrückt als Anzahl der Basenpaare oder Aminosäuren
<212>	Art	O	Art des in SEQ ID NO: x sequenzierten Moleküls, und zwar entweder DNA, RNA oder PRT; enthält eine Nucleotidsequenz

			sowohl DNA- als auch RNA-Fragmente, so ist "DNA" anzugeben; zusätzlich ist das kombinierte DNA-/RNA-Molekül im Merkmalsteil unter <220> bis <223> näher zu beschreiben.
<213>	Organismus	O	Gattung/Art (d. h. wissenschaftlicher Name), "künstliche Sequenz" oder "unbekannt"
<220>	Merkmal	O in den Fällen nach Nr. 29 und 30 Standard	Freilassen; siehe Nr. 29 und 30 Standard; Beschreibung biologisch signifikanter Stellen in der Sequenz gemäß SEQ ID NO: x) (kann je nach der Zahl der angegebenen Merkmale mehrmals vorkommen).
<221>	Name/Schlüssel	O in den Fällen nach Nr. 29 Standard	Siehe Nr. 29 Standard; es dürfen nur die in Anhang 2, Tabelle 5 oder 6 beschriebenen Schlüssel verwendet werden.
<222>	Lage	O in den Fällen nach Nr. 29 Standard	Siehe Nr. 29 Standard; <ul style="list-style-type: none"> <li>- von (Nummer der ersten Base/Aminosäure des Merkmals)</li> <li>- bis (Nummer der letzten Base/Aminosäure des Merkmals)</li> <li>- Basenpaare (Ziffern verweisen auf die Positionen der Basenpaare in einer Nucleotidsequenz)</li> <li>- Aminosäuren (Ziffern verweisen auf die Positionen der Aminosäurereste in einer Aminosäuresequenz)</li> <li>- Angabe, ob sich das Merkmal auf dem zum Strang des Sequenzprotokolls komplementären Strang befindet</li> </ul>
<223>	Sonstige Angaben	O in den Fällen nach Nr. 29 und 30 Standard	Siehe Nr. 29 und 30 Standard; sonstige relevante Angaben, wobei sprachneutrales Vokabular oder freier Text (möglichst in englischer Sprache) zu verwenden ist; freier Text ist im Hauptteil der Beschreibung in der dort verwendeten Sprache zu wiederholen (s. Nr. 36 Standard); enthält die Sequenz eine der in Anhang 2, Tabellen 2 und 4 aufgeführten modifizierten Basen oder modifizierten/seltenen L-Aminosäuren, so ist für diese Base oder Aminosäure das dazugehörige Symbol aus Anhang 2, Tabellen 2 und 4 zu verwenden.
<300>	Veröffentlichungsangaben	F	Freilassen; dieser Abschnitt ist für jede relevante Veröffentlichung zu wiederholen.
<301>	Verfasser	F	
<302>	Titel	F	Titel der Veröffentlichung
<303>	Zeitschrift	F	Name der Zeitschrift, in der die Daten veröffentlicht wurden
<304>	Band	F	Band der Zeitschrift, in dem die Daten veröffentlicht wurden
<305>	Heft	F	Nummer des Hefts der Zeitschrift, in dem die Daten veröffentlicht wurden
<306>	Seiten	F	Seiten der Zeitschrift, auf denen die Daten veröffentlicht wurden



<307>	Datum	F	Datum der Zeitschrift, an dem die Daten veröffentlicht wurden; Angabe nach Möglichkeit entsprechend dem WIPO-Standard ST. 2 (CCYY MM DD)
<308>	Eingangsnummer in der Datenbank	F	von der Datenbank zugeteilte Eingangsnummer einschließlich Datenbankbezeichnung
<309>	Datenbank-Eingabedatum	F	Datum der Eingabe in die Datenbank; Angabe entsprechend dem WIPO-Standard ST. 2 (CCYY MM DD)
<310>	Dokumentennummer	F	Nummer des Dokuments, nur bei Patentdokumenten; die vollständige Nummer hat nacheinander folgendes zu enthalten: den zweibuchstabigen Code entsprechend dem WIPO-Standard ST. 3, die Veröffentlichungsnummer entsprechend dem WIPO-Standard ST. 6 und den Code für die Dokumentenart nach dem WIPO-Standard ST. 16
<311>	Anmeldetag	F	Anmeldetag des Dokuments, nur bei Patentdokumenten; Angabe entsprechend WIPO-Standard ST. 2 (CCYY MM DD)
<312>	Veröffentlichungsdatum	F	Datum der Veröffentlichung des Dokuments; nur bei Patentdokumenten; Angabe entsprechend WIPO-Standard ST. 2 (CCYY MM DD)
<313>	Relevante Reste in SEQ ID NO: x von bis	F	
<400>	Sequenz	O	SEQ ID NO: x sollte in der der Sequenz vorausgehenden Zeile hinter der numerischen Kennzahl stehen (s. Anlage 3).

**Anhang 2****Symbole für Nucleotide und Aminosäuren und Merkmalstabellen**

Tabelle 1: Liste der Nucleotide

Symbol	Bedeutung	Ableitung der Bezeichnung
a	a	<u>A</u> denin
g	g	<u>G</u> uanin
c	c	<u>C</u> ytosin
t	t	<u>T</u> hymine
u	u	<u>U</u> racil
r	g oder a	Pu <u>r</u> in
y	t/u oder c	Pyri <u>m</u> idin
m	a oder c	A <u>m</u> ino
k	g oder t/u	<u>K</u> eto
s	g oder c	<u>s</u> tarke Bindungen 3 H-Brücken
w	a oder t/u	schwache (e: <u>w</u> weak) Bindungen 2 H-Brücken
b	b oder c oder t/u	nicht a
d	a oder g oder t/u	nicht c
h	a oder c oder t/u	nicht g
v	a oder g oder c	nicht t, nicht u
n	a oder g oder c oder t/u, unbekannt oder sonstige	beliebig (e: <u>a</u> ny)

Tabelle 2: Liste der modifizierten Nucleotide

Symbol	Bedeutung
ac4c	4-Acetylcytidin
chm5u	5-(Carboxyhydroxymethyl)uridin
cm	2'-O-Methylcytidin
cmnm5s2u	5-Carboxymethylaminomethyl-2-thiouridin
cmnm5u	5-Carboxymethylaminomethyluridin
d	Dihydrouridin
fm	2'-O-Methylpseudouridin
gal q	beta, D-Galactosylqueuosin
gm	2'-O-Methylguanosin
i	Inosin
i6a	N6-Isopentenyladenosin
m1a	1-Methyladenosin
m1f	1-Methylpseudouridin
m1g	1-Methylguanosin
m1i	1-Methylinosin
m22g	2,2-Dimethylguanosin
m2a	2-Methyladenosin
m2g	2-Methylguanosin
m3c	3-Methylcytidin
m5c	5-Methylcytidin
m6a	N6-Methyladenosin
m7g	7-Methylguanosin
mam5u	5-Methylaminomethyluridin
mam5s2u	5-Methoxyaminomethyl-2-thiouridin
man q	beta, D-Mannosequeuosin
mcm5s2u	5-Methoxycarbonylmethyl-2-thiouridin
mcm5u	5-Methoxycarbonylmethyluridin
mo5u	5-Methoxyuridin
ms2i6a	2-Methylthio-N6-isopentenyladenosin
ms2t6a	N-((9-beta-D-Ribofuranosyl-2-methylthiopurin-6-yl) carbamoyl) threonin
mt6a	N-((9-beta-D-Ribofuranosylpurin-6-yl)N-methylcarbamoyl)threonin
mv	Uridin-5-oxyessigsäuremethylester
o5u	Uridin-5-oxyessigsäure (v)
osyw	Wybutoxosin
p	Pseudouridin

q	Qeuosin
s2c	2-Thiocytidin
s2t	5-Methyl-2-thiouridin
s2u	2-Thiouridin
s4u	4-Thiouridin
t	5-Methyluridin
t6a	N-((9-beta-D-Ribofuranosylpurin-6-yl) carbamoyl)threonin
tm	2'-O-Methyl-5-methyluridin
um	2'-O-Methyluridin
yw	Wybutosin
x	3-(3-Amino-3-carboxypropyl)uridin, (acp3)u

Tabelle 3: Liste der Aminosäuren

Symbol	Bedeutung
Ala	Alanin
Cys	Cystein
Asp	Asparaginsäure
Glu	Glutaminsäure
Phe	Phenylalanin
Gly	Glycin
His	Histidin
Ile	Isoleucin
Lys	Lysin
Leu	Leucin
Met	Methionin
Asn	Asparagin
Pro	Prolin
Gln	Glutamin
Arg	Arginin
Ser	Serin
Thr	Threonin
Val	Valin
Trp	Tryptophan
Tyr	Tyrosin
Asx	Asp oder Asn
Glx	Glu oder Gln
Xaa	unbekannt oder sonstige

Tabelle 4: Liste der modifizierten und seltenen Aminosäuren

Symbol	Bedeutung
Aad	2-Aminoadipinsäure
bAad	3-Aminoadipinsäure
bAla	beta-Alanin, beta-Aminopropionsäure
Abu	2-Aminobuttersäure
4Abu	4-Aminobuttersäure, Piperidinsäure
Acp	6-Aminocaprinsäure
Ahe	2-Aminoheptansäure
Aib	2-Aminoisobuttersäure
bAib	3-Aminoisobuttersäure
Apm	2-Aminopimelinsäure
Dbu	2,4-Diaminobuttersäure
Des	Desmosin
Dpm	2,2'-Diaminopimelinsäure
Dpr	2,3-Diaminopropionsäure
EtGly	N-Ethylglycin
EtAsn	N-Ethylasparagin
Hyl	Hydroxylysin
aHyl	allo-Hydroxylysin
3Hyp	3-Hydroxyprolin
4Hyp	4-Hydroxyprolin
Ide	Isodesmosin
alle	allo-Isoleucin
MeGly	N-Methylglycin, Sarkosin
Melle	N-Methylisoleucin
MeLys	6-N-Methyllysin
MeVal	N-Methylvalin
Nva	Norvalin
Nle	Norleucin
Orn	Ornithin

Tabelle 5: Liste der Merkmalschlüssel zu Nucleotidsequenzen

Schlüssel	Beschreibung
allele	Ein verwandtes Individuum oder ein verwandter Stamm enthält stabile alternative Formen desselben Gens und unterscheidet sich an dieser (und vielleicht an anderer) Stelle von der vorliegenden Sequenz.
attenuator	1. Region einer DNA, in der die Beendigung der Transkription reguliert wird und die Expression einiger bakterieller Operons gesteuert wird 2. zwischen dem Promotor und dem ersten Strukturgen liegender Sequenzabschnitt, der eine partielle Beendigung der Transkription bewirkt
C_region	konstante Region der leichten und schweren Immunglobulinketten und der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; enthält je nach Kette ein oder mehrere Exons
CAAT_signal	CAAT-Box; Teil einer konservierten Sequenz, der etwa 75 Basenpaare stromaufwärts vom Startpunkt der eukaryontischen Transkriptionseinheiten liegt und an der RNA-Polymerase-Bindung beteiligt sein kann; Konsensussequenz = GG (C oder T) CAATCT
CDS	codierende Sequenz; Sequenz von Nucleotiden, die mit der Sequenz der Aminosäuren in einem Protein übereinstimmt (beinhaltet Stopcodon); Merkmal schließt eine mögliche Translation der Aminosäure ein
conflict	Unabhängige Determinierungen "derselben" Sequenz unterscheiden sich an dieser Stelle oder in dieser Region voneinander.
D-loop	D-Schleife; Region innerhalb der mitochondrialen DNA, in der sich ein kürzeres RNA-Stück mit einem Strang der doppelsträngigen DNA paart und dabei den ursprünglichen Schwesterstrang in dieser Region verdrängt; dient auch zur Beschreibung der Verdrängung einer einzelsträngigen Region des DNA-Doppelstrangs durch einen einzelsträngigen Eindringling bei der durch ein recA-Protein ausgelösten Reaktion
D-segment	Diversity-Region der schweren Kette von Immunglobulin und der Beta-Kette eines T-Zell-Rezeptors
enhancer	eine als Cis-Element wirkende Sequenz, die die Aktivität (einiger) eukaryontischer Promotoren verstärkt und in beliebiger Richtung und Position zum Promotor (stromaufwärts oder -abwärts) funktioniert
exon	Region des Genoms, die für einen Teil der gespleißten mRNA codiert; kann 5'UTR, alle CDSs und 3'UTR enthalten
GC_signal	GC-Box; eine konservierte, GC-reiche Region stromaufwärts vom Startpunkt der eukaryontischen Transkriptionseinheiten, die in mehreren Kopien und in beiden Richtungen vorkommen kann; Konsensussequenz = GGGCGG
gene	biologisch signifikante Region, die als Gen bezeichnet wird und einen Namen trägt
iDNA	intervenierende DNA; DNA, die durch verschiedene Arten der Rekombination eliminiert wird
intron	DNA-Abschnitt, der transkribiert, aber beim Zusammenspleißen der ihn umrahmenden Sequenzen (Exons) aus dem Transkript wieder herausgeschnitten wird
J_segment	J-Kette (Verbindungskette) zwischen den leichten und den schweren Immunglobulinketten und den Alpha-, Beta- und Gamma-Ketten der T-Zell-Rezeptoren
LTR	lange Terminalwiederholung, eine an den beiden Enden einer gegebenen Sequenz direkt wiederholte Sequenz, wie sie für Retroviren typisch ist
mat_peptide	für ein reifes Peptid oder Protein codierende Sequenz; Sequenz, die im Anschluß an eine posttranslationale Modifizierung für das reife oder endgültige Peptid- oder Proteinprodukt codiert; schließt im Gegensatz zur entsprechenden CDS das Stopcodon nicht ein
misc_binding	Stelle in einer Nucleinsäure, die einen anderen Teil, der nicht durch einen anderen Bindungsschlüssel (primer_bind oder protein_bind) beschrieben werden kann, kovalent oder nicht kovalent bindet
misc_difference	Die Merkmalsequenz unterscheidet sich von der im Eintrag und kann nicht durch einen anderen Unterscheidungsschlüssel (conflict, unsure, old_sequence, mutation, variation, allele bzw. modified_base) beschrieben werden.

misc_feature	biologisch signifikante Region, die nicht durch einen anderen Merkmalschlüssel beschrieben werden kann; neues oder seltenes Merkmal
misc_recomb	Stelle, an der ein allgemeiner, ortsspezifischer oder replikativer Rekombinationsvorgang stattfindet, bei dem die DNA-Doppelhelix aufgebrochen und wieder zusammengefügt wird, und die nicht durch andere Rekombinationsschlüssel (iDNA oder Virion) oder den betreffenden Herkunftsschlüssel (/insertions_seq, /transposon, /proviral) beschrieben werden kann
misc_RNA	Transkript oder RNA-Produkt, das nicht durch andere RNA-Schlüssel (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 oder snRNA) beschrieben werden kann
misc_signal	Region, die ein Signal enthält, das die Genfunktion oder -expression steuert oder ändert, und die nicht durch andere Signalschlüssel (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator oder rep_origin) beschrieben werden kann
misc_structure	Sekundär-, Tertiär- oder sonstige Struktur oder Konformation, die nicht durch andere Strukturschlüssel (stem_loop oder D-loop) beschrieben werden kann
modified_base	Das angegebene Nucleotid ist ein modifiziertes Nucleotid und ist durch das angegebene Molekül (ausgedrückt durch die modifizierte Base) zu ersetzen.
mRNA	messenger-RNA (Boten-RNA); enthält eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon) und eine 3'-nichttranslatierte Region (3'UTR)
mutation	Ein verwandter Stamm weist an dieser Stelle eine plötzliche, erbliche Sequenzveränderung auf.
N_region	zusätzliche Nucleotide, die zwischen neu geordnete Immunglobulinabschnitte eingefügt werden
old_sequence	Die vorliegende Sequenz stellt eine geänderte Version der früher an dieser Stelle befindlichen Sequenz dar.
polyA_signal	Erkennungsregion, die zur Endonuclease-Spaltung eines RNA-Transkripts mit anschließender Polyadenylierung nötig ist; Konsensussequenz: AATAAA
polyA_site	Stelle auf einem RNA-Transkript, an der durch posttranslationale Polyadenylierung Adeninreste eingefügt werden
precursor_RNA	noch nicht gereifte RNA-Spezies; kann eine am 5'-Ende abzuschneidende Region (5'Clip), eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron), eine 3'-nichttranslatierte Region (3'UTR) und eine am 3'-Ende abzuschneidende Region (3'Clip) enthalten
prim_transcript	primäres (ursprüngliches, nicht prozessiertes) Transkript; enthält eine am 5'-Ende abzuschneidende Region (5'Clip), eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron), eine 3'-nichttranslatierte Region (3'UTR) und eine am 3'-Ende abzuschneidende Region (3'Clip)
primer_bind	nichtkovalente Primer-Bindungsstelle für die Initiierung der Replikation, Transkription oder reversen Transkription; enthält eine oder mehrere Stellen für synthetische Elemente, z. B. PCR-Primerelemente
promoter	Region auf einem DNA-Molekül, die an der Bindung der RNA-Polymerase beteiligt ist, die die Transkription initiiert
protein_bind	nichtkovalente Protein-Bindungsstelle auf der Nucleinsäure
RBS	ribosomale Bindungsstelle
repeat_region	Genomregion mit repetitiven Einheiten
repeat_unit	einzelnes Repeat (repetitive Einheit)
rep_origin	Replikationsursprung; Startpunkt der Duplikation der Nucleinsäure, durch die zwei identische Kopien entstehen
rRNA	reife ribosomale RNA; RNA-Komplex des Ribonucleoprotein-Partikels (Ribosom), der Aminosäuren zu Proteinen zusammenfügt
S_region	Switch-Region der schweren Immunglobulinketten; beteiligt am Umbau schwerer DNA-Ketten, der zur Expression einer anderen Immunglobulin-Klasse aus derselben B-Zelle führt

satellite	viele tandemartig hintereinandergeschaltete (identische oder verwandte) Repeats einer kurzen basischen repetitiven Einheit; viele davon unterscheiden sich in der Basenzusammensetzung oder einer anderen Eigenschaft vom Genomdurchschnitt und können so von der Hauptmasse der genomischen DNA abgetrennt werden
scRNA	kleine cytoplasmische RNA; eines von mehreren kleinen cytoplasmischen RNA-Molekülen im Cytoplasma und (manchmal) im Zellkern eines Eukaryonten
sig_peptide	für ein Signalpeptid codierende Sequenz; Sequenz, die für eine N-terminale Domäne eines sekretorischen Proteins codiert; diese Domäne spielt bei der Anheftung des naszierenden Polypeptids an die Membran eine Rolle; Leader-Sequenz
snRNA	kleine Kern-RNA; eine der vielen kleinen RNA-Formen, die nur im Zellkern vorkommen; einige der snRNAs spielen beim Spleißen oder bei anderen RNA-verarbeitenden Reaktionen eine Rolle
source	Bezeichnet die biologische Herkunft des genannten Sequenzabschnitts; die Angabe dieses Schlüssels ist obligatorisch; jeder Eintrag muß mindestens einen Herkunftsschlüssel aufweisen, der die gesamte Sequenz umfaßt; es dürfen zu jeder Sequenz auch mehrere Herkunftsschlüssel angegeben werden.
stem_loop	Haarnadelschleife; eine Doppelhelix-Region, die durch Basenpaarung zwischen benachbarten (invertierten) komplementären Sequenzen in einem RNA- oder DNA-Einzelstrang entsteht
STS	Sequence Tagged Site; kurze, nur als Einzelkopie vorkommende DNA-Sequenz, die einen Kartierungspunkt auf dem Genom bezeichnet und durch PCR ermittelt werden kann; eine Region auf dem Genom kann durch Bestimmung der Reihenfolge der STSs kartiert werden
TATA_signal	TATA-Box; Goldberg-Hogness-Box; ein konserviertes AT-reiches Septamer, das sich rund 25 Basenpaare vor dem Startpunkt jeder eukaryontischen RNA-Polymerase-II-Transkriptionseinheit befindet und bei der Positionierung des Enzyms für eine korrekte Initiation eine Rolle spielen kann; Konsensussequenz = TATA(A oder T)A(A oder T)
terminator	DNA-Sequenz, die entweder am Ende des Transkripts oder neben einer Promotor-Region liegt und bewirkt, daß die RNA-Polymerase die Transkription beendet; kann auch die Bindungsstelle eines Repressor-Proteins sein
transit_peptide	für Transitpeptid codierende Sequenz; Sequenz, die für eine N-terminale Domäne eines im Zellkern codierten Organellen-Proteins codiert; diese Domäne ist an der posttranslationalen Einschleusung des Proteins in die Organelle beteiligt
tRNA	reife transfer-RNA, ein kleines RNA-Molekül (75–85 Basen lang), das die Translation einer Nucleinsäure-Sequenz in eine Aminosäure-Sequenz vermittelt
unsure	Der Autor kennt die Sequenz in dieser Region nicht genau.
V_region	variable Region der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; codiert für das variable Aminoende; kann aus V-, D-, N- und J-Abschnitten bestehen
V-segment	variabler Abschnitt der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; codiert für den Großteil der variablen Region (V_region) und die letzten Aminosäuren des Leader-Peptids
variation	Ein verwandter Stamm enthält stabile Mutationen desselben Gens (z. B. RFLPs, Polymorphismen usw.), die sich an dieser (und möglicherweise auch an anderer) Stelle von der vorliegenden Sequenz unterscheiden.
3'clip	3'-äußerste Region eines Precursor-Transkripts, die bei der Prozessierung abgeschnitten wird
3'UTR	Region am 3'-Ende eines reifen Transkripts (nach dem Stopcodon), die nicht in ein Protein translatiert wird
5'clip	5'-äußerste Region eines Precursor-Transkripts, die bei der Prozessierung abgeschnitten wird
5'UTR	Region am 5'-Ende eines reifen Transkripts (vor dem Initiationscodon), die nicht in ein Protein translatiert wird
-10_signal	Pribnow-Box; konservierte Region rund 10 Basenpaare stromaufwärts vom Startpunkt der bakteriellen Transkriptionseinheiten, die bei der Bindung der RNA-Polymerase eine Rolle spielt; Konsensussequenz = TAtAAT
-35_signal	konserviertes Hexamer rund 35 Basenpaare stromaufwärts vom Startpunkt der bakteriellen Transkriptionseinheiten; Konsensussequenz = TTGACa [ ] oder TGTTGACA [ ]



Tabelle 6: Liste der Merkmalschlüssel zu Proteinsequenzen

Schlüssel	Beschreibung
CONFLICT	In den einzelnen Unterlagen ist von verschiedenen Sequenzen die Rede.
VARIANT	Den Angaben der Autoren zufolge gibt es Sequenzvarianten.
VARSPLIC	Beschreibung von Sequenzvarianten, die durch alternatives Spleißen entstanden sind
MUTAGEN	experimentell veränderte Stelle
MOD_RES	posttranslationale Modifikation eines Rests
ACETYLATION	N-terminale oder sonstige
AMIDATION	in der Regel am C-Terminus eines reifen aktiven Peptids
BLOCKED	unbestimmte Gruppe, die das N- oder C-terminale Ende blockiert
FORMYLATION	des N-terminalen Methionin
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	von Asparagin, Asparaginsäure, Prolin oder Lysin
METHYLATION	in der Regel von Lysin oder Arginin
PHOSORYLATION	von Serin, Threonin, Tyrosin, Asparaginsäure oder Histidin
PYRROLIDONE	N-terminales Glutamat, das ein internes cyclisches Lactam gebildet hat
CARBOXYLIC ACID SULFATATION	in der Regel von Tyrosin
LIPID	kovalente Bindung einer Lipidhälfte
MYRISTATE	Myristat-Gruppe, die durch eine Amidbindung an den N-terminalen Glycin-Rest der reifen Form eines Proteins oder an einen internen Lysin-Rest gebunden ist
PALMITATE	Palmitat-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest oder durch eine Esterbindung an einen Serin- oder Threonin-Rest gebunden ist
FARNESYL	Farnesyl-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest gebunden ist
GERANYL-GERANYL	Geranyl-geranyl-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest gebunden ist
GPI-ANCHOR	Glykosyl-phosphatidylinositol-(GPI-)Gruppe, die an die alpha-Carboxylgruppe des C-terminalen Rests der reifen Form eines Proteins gebunden ist
N-ACYL DIGLYCERIDE	N-terminales Cystein der reifen Form eines prokaryontischen Lipoproteins mit einer amidgebundenen Fettsäure und einer Glyceryl-Gruppe, an die durch Esterbindungen zwei Fettsäuren gebunden sind
DISULFID	Disulfidbindung; den "VON"- und den "BIS"-Endpunkt bilden die beiden Reste, die durch eine ketteninterne Disulfidbindung verbunden sind; sind der "VON"- und der "BIS"-Endpunkt identisch, liegt die Disulfidbindung innerhalb der Kette, und die Art der Quervernetzung ist im Beschreibungsfeld anzugeben
THIOLEST	Thiolesterbindung; den "VON"- und den "BIS"-Endpunkt bilden die beiden Reste, die durch die Thiolesterbindung verbunden sind
THIOETH	Thioetherbindung; den "VON"- und den "BIS"-Endpunkt bilden die beiden Reste, die durch die Thioetherbindung verbunden sind
CARBOHYD	Glykosylierungs-Stelle; die Art des Kohlenhydrats (sofern bekannt) ist im Beschreibungsfeld anzugeben
METAL	Bindungsstelle für ein Metallion; die Art des Metalls ist im Beschreibungsfeld anzugeben

BINDING	Bindungsstelle für eine beliebige chemische Gruppe (Coenzym, prosthetische Gruppe usw.); die Art der Gruppe ist im Beschreibungsfeld anzugeben
SIGNAL	Bereich einer Signalsequenz (Präpeptid)
TRANSIT	Bereich eines Transit-Peptids (mitochondriales, chloroplastisches oder für Microbodies)
PROPEP	Bereich eines Propeptids
CHAIN	Bereich einer Polypeptid-Kette im reifen Protein
PEPTIDE	Bereich eines freigesetzten aktiven Peptids
DOMAIN	Bereich einer wichtigen Domäne auf der Sequenz; die Art dieser Domäne ist im Beschreibungsfeld anzugeben
CA_BIND	Bereich einer Calcium-bindenden Region
DNA_BIND	Bereich einer DNA-bindenden Region
NP_BIND	Bereich einer Nucleotidphosphat-bindenden Region; die Art des Nucleotidphosphats ist im Beschreibungsfeld anzugeben
TRANSMEM	Bereich einer Transmembran-Region
ZN_FING	Bereich einer Zink-Finger-Region
SIMILAR	Grad der Ähnlichkeit mit einer anderen Proteinsequenz; im Beschreibungsfeld sind genaue Angaben über diese Sequenz zu machen
REPEAT	Bereich einer internen Sequenzwiederholung
HELIX	Sekundärstruktur: Helices, z. B. Alpha-Helix, 3(10)-Helix oder Pi-Helix
STRAND	Sekundärstruktur: Beta-Strang, z. B. durch Wasserstoff-Brückenbindungen stabilisierter Beta-Strang, oder Rest in einer isolierten Beta-Brücke
TURN	Sekundärstruktur: Schleife, z. B. durch Wasserstoff-Brückenbindungen stabilisierte Schleife (3-, 4- oder 5-Schleife)
ACT_SITE	Aminosäure(n), die bei der Aktivität eines Enzyms mitwirkt (mitwirken)
SITE	irgendeine andere wichtige Stelle auf der Sequenz
INIT_MET	die Sequenz beginnt bekanntermaßen mit einem Start-Methionin
NON_TER	Der Rest am Sequenzanfang oder -ende ist nicht der Terminalrest; steht er an der Position 1, so bedeutet das, daß diese nicht der N-Terminus des vollständigen Moleküls ist; steht er an letzter Position, so ist diese Position nicht der C-Terminus des vollständigen Moleküls; für diesen Schlüssel gibt es kein Beschreibungsfeld.
NON_CONS	nicht aufeinanderfolgende Reste; zeigt an, daß zwei Reste in einer Sequenz nicht aufeinanderfolgen, sondern daß zwischen ihnen einige nichtsequenzierte Reste liegen
UNSURE	Unsicherheiten in der Sequenz; mit diesem Schlüssel werden Regionen einer Sequenz beschrieben, bei der sich der Autor bezüglich der Sequenzzuweisung nicht sicher ist

### Anhang 3

#### Muster eines Sequenzprotokolls

```

<110>      Smith, John; Smithgene Inc.
<120>      Beispiel für ein Sequenzprotokoll
<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  cgcgctcctc  tcgcgcctct  ctctcgctct  cctctcgctc          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
    
```

<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> Künstliche Sequenz

<220>  
 <223> Peptid, das nach Größe und Polarität als Verbindung zwischen den Alpha- und Beta-Ketten des Proteins XYZ konzipiert ist

<400> 3

Met	Val	Asn	Leu	Glu	Pro	Met	His	Thr	Glu	Ile
1				5					10	

<210> 4  
 <400> 4  
 000

## Appendix 1

### Numeric identifiers

Only numeric identifiers as defined below may be used in sequence listings submitted in applications. The text of the data element headings given below shall not be included in the sequence listings.

Numeric identifiers of mandatory data elements, that is, data elements which must be included in all sequence listings (see paragraph 25 of this Standard: items 110, 120, 160, 210, 211, 212, 213 and 400) and numeric identifiers of data elements which must be included in circumstances specified in this Standard (see paragraphs 26, 27, 28, 29 and 30

of this Standard: items 130, 140, 141, 150 and 151, and 220 to 223) are marked by the symbol "M".

Numeric identifiers of optional data elements (see paragraph 31 of this Standard) are marked by the symbol "O".

Numeric Identifier	Numeric Identifier Description	Mandatory (M) or Optional (O)	Comment
<110>	Applicant name	M	where the name of the applicant is written in characters other than those of the Latin alphabet, the same shall also be indicated in characters of the Latin alphabet either as a mere transliteration or through translation into English
<120>	Title of invention	M	
<130>	File reference	M, in the circumstances specified in paragraph 26 of this Standard	see paragraph 26 of this Standard
<140>	Current patent application	M, in the circumstances specified in paragraph 27 of this Standard	see paragraph 27 of this Standard; the current patent application shall be identified, in the following order, by the two-letter code indicated in accordance with WIPO Standard ST.3 and the application number (in the format used by the industrial property Office with which the current patent application is filed) or, for an international application, by the international application number
<141>	Current filing date	M, in the circumstances specified in paragraph 27 of this Standard	see paragraph 27 of this Standard; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<150>	Earlier patent application	M, in the circumstances specified in paragraph 28 of this Standard	see paragraph 28 of this Standard; the earlier patent application shall be identified, in the following order, by the two-letter code indicated in accordance with WIPO Standard ST.3 and the application number (in the format used by the industrial property Office with which the earlier patent application was filed) or, for an international application, by the international application number
<151>	Earlier application filing date	M, in the circumstances specified in paragraph 28 of this Standard	see paragraph 28 of this Standard; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<160>	Number of SEQ ID NOs	M	
<170>	Software	O	
<210>	Information for SEQ ID NO: x	M	response shall be an integer representing the SEQ ID NO shown
<211>	Length	M	sequence length expressed in number of base pairs or amino acids
<212>	Type	M	type of molecule sequenced in SEQ ID NO: x, either DNA, RNA or PRT; if a nucleotide sequence contains both DNA

			and RNA fragments, the value shall be "DNA"; in addition, the combined DNA/RNA molecule shall be further described in the <220> to <223> feature section
<213>	Organism	M	Genus Species (that is, scientific name) or "Artificial Sequence" or "Unknown"
<220>	Feature	M, in the circumstances specified in paragraph 29 and 30 of this Standard	leave blank; see paragraphs 29 and 30 of this Standard; description of points of biological significance in the sequence in SEQ ID NO: x) (may be repeated depending on the number of features indicated)
<221>	Name/key	M, in the circumstances specified in paragraph 29 of this Standard	see paragraph 29 of this Standard; only those keys as described in Table 5 or 6 of Appendix 2 shall be used
<222>	Location	M, in the circumstances specified in paragraph 29 of this Standard	see paragraph 29 of this Standard; <ul style="list-style-type: none"> <li>- from (number of first base/amino acid in the feature)</li> <li>- to (number of last base/amino acid in the feature)</li> <li>- base pairs (numbers refer to positions of base pairs in a nucleotide sequence)</li> <li>- amino acids (numbers refer to positions of amino acid residues in an amino acid sequence)</li> <li>- whether feature is located on the complementary strand to that filed in the sequence listing</li> </ul>
<223>	Other information	M, in the circumstances specified in paragraphs 29 and 30 of this Standard	see paragraphs 29 and 30 of this Standard; any other relevant information, using language neutral vocabulary, or free text (preferably in English); any free text is to be repeated in the main part of the description in the language thereof (see paragraph 36 of this Standard); where any modified base or modified/unusual L-amino acid appearing in Appendix 2, Tables 2 and 4, is in the sequence, the symbol associated with that base or amino acid from Appendix 2, Tables 2 and 4, should be used
<300>	Publication information	O	leave blank; repeat section for each relevant publication
<301>	Authors	O	
<302>	Title	O	title of publication
<303>	Journal	O	journal name in which data published
<304>	Volume	O	journal volume in which data published
<305>	Issue	O	journal issue number in which data published
<306>	Pages	O	journal page numbers on which data published
<307>	Date	O	journal date on which data published; if possible, the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<308>	Database accession number	O	accession number assigned by database including database name

<309>	Database entry date	O	date of entry in database; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<310>	Document number	O	document number, for patent type citations only; the full document shall specify, in the following order, the two-letter code indicated in accordance with WIPO Standard ST.3, the publication number indicated in accordance with WIPO Standard ST.6, and the kind-of-document code indicated in accordance with WIPO Standard ST.16
<311>	Filing date	O	document filing date, for patent-type citations only; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<312>	Publication date	O	document publication date; for patent-type citations only; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<313>	Relevant residues in SEQ ID NO: x: from to	O	
<400>	Sequence	M	SEQ ID NO: x should follow the numeric identifier and should appear on the line preceding the sequence (see Appendix 3)

## Appendix 2

### Nucleotide and amino acid symbols and feature table

Table 1: List of Nucleotides

Symbol	Meaning	Origin of designation
a	a	<u>a</u> denine
g	g	g <u>u</u> anine
c	c	<u>c</u> ytosine
t	t	<u>t</u> hymine
u	u	<u>u</u> racil
r	g or a	pur <u>r</u> ine
y	t/u or c	pyr <u>y</u> midine
m	a or c	<u>a</u> mino
k	g or t/u	<u>k</u> eto
s	g or c	<u>s</u> trong interactions 3H-bonds
w	a or t/u	<u>w</u> weak interactions 2H-bonds
b	g or c or t/u	not a
d	a or g or t/u	not c
h	a or c or t/u	not g
v	a or g or c	not t, not u
n	a or g or c or t/u, unknown, or other	<u>a</u> ny



Table 2: List of Modified Nucleotides

Symbol	Meaning
ac4c	4-acetylcytidine
chm5u	5-(carboxyhydroxymethyl)uridine
cm	2'-O-methylcytidine
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine
cmnm5u	5-carboxymethylaminomethyluridine
d	dihydrouridine
fm	2'-O-methylpseudouridine
gal q	beta, D-galactosylqueuosine
gm	2'-O-methylguanosine
i	inosine
i6a	N6-isopentenyladenosine
m1a	1-methyladenosine
m1f	1-methylpseudouridine
m1g	1-methylguanosine
m1i	1-methylinosine
m22g	2,2-dimethylguanosine
m2a	2-methyladenosine
m2g	2-methylguanosine
m3c	3-methylcytidine
m5c	5-methylcytidine
m6a	N6-methyladenosine
m7g	7-methylguanosine
mam5u	5-methylaminomethyluridine
mam5s2u	5-methoxyaminomethyl-2-thiouridine
man q	beta, D-mannosylqueuosine
mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
mcm5u	5-methoxycarbonylmethyluridine
mo5u	5-methoxyuridine
ms2i6a	2-methylthio-N6-isopentenyladenosine
ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methylcarbamoyl)threonine
mv	uridine-5-oxyacetic acid-methylester
o5u	uridine-5-oxyacetic acid
osyw	wybutoxosine
p	pseudouridine

q	queuosine
s2c	2-thiocytidine
s2t	5-methyl-2-thiouridine
s2u	2-thiouridine
s4u	4-thiouridine
t	5-methyluridine
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)-carbamoyl)threonine
tm	2'-O-methyl-5-methyluridine
um	2'-O-methyluridine
yw	wybutosine
x	3-(3-amino-3-carboxy-propyl)uridine, (acp3)u

Table 3: List of Amino Acids

Symbol	Meaning
Ala	Alanine
Cys	Cysteine
Asp	Aspartic Acid
Glu	Glutamic Acid
Phe	Phenylalanine
Gly	Glycine
His	Histidine
Ile	Isoleucine
Lys	Lysine
Leu	Leucine
Met	Methionine
Asn	Asparagine
Pro	Proline
Gln	Glutamine
Arg	Arginine
Ser	Serine
Thr	Threonine
Val	Valine
Trp	Tryptophan
Tyr	Tyrosine
Asx	Asp or Asn
Glx	Glu or Gln
Xaa	unknown or other

**Table 4: List of Modified and Unusual Amino Acids**

Symbol	Meaning
Aad	2-Aminoadipic acid
bAad	3-Aminoadipic acid
bAla	beta-Alanine, beta-Aminopropionic acid
Abu	2-Aminobutyric acid
4Abu	4-Aminobutyric acid, piperidinic acid
Acp	6-Aminocaproic acid
Ahe	2-Aminoheptanoic acid
Aib	2-Aminoisobutyric acid
bAib	3-Aminoisobutyric acid
Apm	2-Aminopimelic acid
Dbu	2,4-Diaminobutyric acid
Des	Desmosine
Dpm	2,2'-Diaminopimelic acid
Dpr	2,3-Diaminopropionic acid
EtGly	N-Ethylglycine
EtAsn	N-Ethylasparagine
Hyl	Hydroxylysine
aHyl	allo-Hydroxylysine
3Hyp	3-Hydroxyproline
4Hyp	4-Hydroxyproline
Ide	Isodesmosine
alle	allo-Isoleucine
MeGly	N-Methylglycine, sarcosine
Melle	N-Methylisoleucine
MeLys	6-N-Methyllysine
MeVal	N-Methylvaline
Nva	Norvaline
Nle	Norleucine
Orn	Ornithine

**Table 5: List of Feature Keys Related to Nucleotide Sequences**

Key	Description
allele	a related individual or strain contains stable, alternative forms of the same gene which differs from the presented sequence at this location (and perhaps others)
attenuator	(1) region of DNA at which regulation of termination of transcription occurs, which controls the expression of some bacterial operons; (2) sequence segment located between the promoter and the first structural gene that causes partial termination of transcription
C_region	constant region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; includes one or more exons depending on the particular chain
CAAT_signal	CAAT box; part of a conserved sequence located about 75 bp up-stream of the start point of eukaryotic transcription units which may be involved in RNA polymerase binding; consensus=GG (C or T) CAATCT
CDS	coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon); feature includes amino acid conceptual translation
conflict	independent determinations of the "same" sequence differ at this site or region
D-loop	displacement loop; a region within mitochondrial DNA in which a short stretch of RNA is paired with one strand of DNA, displacing the original partner DNA strand in this region; also used to describe the displacement of a region of one strand of duplex DNA by a single stranded invader in the reaction catalyzed by RecA protein
D-segment	diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain
enhancer	a cis-acting sequence that increases the utilization of (some) eukaryotic promoters, and can function in either orientation and in any location (upstream or downstream) relative to the promoter
exon	region of genome that codes for portion of spliced mRNA; may contain 5'UTR, all CDSs, and 3'UTR
GC_signal	GC box; a conserved GC-rich region located upstream of the start point of eukaryotic transcription units which may occur in multiple copies or in either orientation; consensus=GGGCGG
gene	region of biological interest identified as a gene and for which a name has been assigned
iDNA	intervening DNA; DNA which is eliminated through any of several kinds of recombination
intron	a segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it
J_segment	joining segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains
LTR	long terminal repeat, a sequence directly repeated at both ends of a defined sequence, of the sort typically found in retroviruses
mat_peptide	mature peptide or protein coding sequence; coding sequence for the mature or final peptide or protein product following post-translational modification; the location does not include the stop codon (unlike the corresponding CDS)
misc_binding	site in nucleic acid which covalently or non-covalently binds another moiety that cannot be described by any other Binding key (primer_bind or protein_bind)
misc_difference	feature sequence is different from that presented in the entry and cannot be described by any other Difference key (conflict, unsure, old_sequence, mutation, variation, allele, or modified_base)
misc_feature	region of biological interest which cannot be described by any other feature key; a new or rare feature
misc_recomb	site of any generalized, site-specific or replicative recombination event where there is a breakage and reunion of duplex DNA that cannot be described by other recombination keys (iDNA and virion) or qualifiers of source key (/insertion_seq, /transposon, /proviral)

misc_RNA	any transcript or RNA product that cannot be defined by other RNA keys (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, and snRNA)
misc_signal	any region containing a signal controlling or altering gene function or expression that cannot be described by other Signal keys (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator, and rep_origin)
misc_structure	any secondary or tertiary structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop)
modified_base	the indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value)
mRNA	messenger RNA; includes 5' untranslated region (5'UTR), coding sequences (CDS, exon) and 3' untranslated region (3'UTR)
mutation	a related strain has an abrupt, inheritable change in the sequence at this location
N_region	extra nucleotides inserted between rearranged immunoglobulin segments
old_sequence	the presented sequence revises a previous version of the sequence at this location
polyA_signal	recognition region necessary for endonuclease cleavage of an RNA transcript that is followed by polyadenylation; consensus=AATAAA
polyA_site	site on an RNA transcript to which will be added adenine residues by post-transcriptional polyadenylation
precursor_RNA	any RNA species that is not yet the mature RNA product; may include 5' clipped region (5'clip), 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron), 3' untranslated region (3'UTR), and 3' clipped region (3'clip)
prim_transcript	primary (initial, unprocessed) transcript; includes 5' clipped region (5'clip), 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron), 3' untranslated region (3'UTR), and 3' clipped region (3'clip)
primer_bind	non-covalent primer binding site for initiation of replication, transcription, or reverse transcription; includes site(s) for synthetic, for example, PCR primer elements
promoter	region on a DNA molecule involved in RNA polymerase binding to initiate transcription
protein_bind	non-covalent protein binding site on nucleic acid
RBS	ribosome binding site
repeat_region	region of genome containing repeating units
repeat_unit	single repeat element
rep_origin	origin of replication; starting site for duplication of nucleic acid to give two identical copies
rRNA	mature ribosomal RNA; the RNA component of the ribonucleoprotein particle (ribosome) which assembles amino acids into proteins
S_region	switch region of immunoglobulin heavy chains; involved in the rearrangement of heavy chain DNA leading to the expression of a different immunoglobulin class from the same B-cell
satellite	many tandem repeats (identical or related) of a short basic repeating unit; many have a base composition or other property different from the genome average that allows them to be separated from the bulk (main band) genomic DNA
scRNA	small cytoplasmic RNA; any one of several small cytoplasmic RNA molecules present in the cytoplasm and (sometimes) nucleus of a eukaryote
sig_peptide	signal peptide coding sequence; coding sequence for an N-terminal domain of a secreted protein; this domain is involved in attaching nascent polypeptide to the membrane; leader sequence
snRNA	small nuclear RNA; any one of many small RNA species confined to the nucleus; several of the snRNAs are involved in splicing or other RNA processing reactions

source	identifies the biological source of the specified span of the sequence; this key is mandatory; every entry will have, as a minimum, a single source key spanning the entire sequence; more than one source key per sequence is permissible
stem_loop	hairpin; a double-helical region formed by base-pairing between adjacent (inverted) complementary sequences in a single strand of RNA or DNA
STS	Sequence Tagged Site; short, single-copy DNA sequence that characterizes a mapping landmark on the genome and can be detected by PCR; a region of the genome can be mapped by determining the order of a series of STSs
TATA_signal	TATA box; Goldberg-Hogness box; a conserved AT-rich septamer found about 25 bp before the start point of each eukaryotic RNA polymerase II transcript unit which may be involved in positioning the enzyme for correct initiation; consensus=TATA(A or T)A(A or T)
terminator	sequence of DNA located either at the end of the transcript or adjacent to a promoter region that causes RNA polymerase to terminate transcription; may also be site of binding of repressor protein
transit_peptide	transit peptide coding sequence; coding sequence for an N-terminal domain of a nuclear-encoded organellar protein; this domain is involved in post-translational import of the protein into the organelle
tRNA	mature transfer RNA, a small RNA molecule (75–85 bases long) that mediates the translation of a nucleic acid sequence into an amino acid sequence
unsure	author is unsure of exact sequence in this region
V_region	variable region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for the variable amino terminal portion; can be made up from V_segments, D_segments, N_regions, and J_segments
V_segment	variable segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for most of the variable region (V_region) and the last few amino acids of the leader peptide
variation	a related strain contains stable mutations from the same gene (for example, RFLPs, polymorphisms, etc.) which differ from the presented sequence at this location (and possibly others)
3'clip	3'-most region of a precursor transcript that is clipped off during processing
3'UTR	region at the 3' end of a mature transcript (following the stop codon) that is not translated into a protein
5'clip	5'-most region of a precursor transcript that is clipped off during processing
5'UTR	region at the 5' end of a mature transcript (preceding the initiation codon) that is not translated into a protein
-10_signal	pribnow box; a conserved region about 10 bp upstream of the start point of bacterial transcription units which may be involved in binding RNA polymerase; consensus=TAtAaT
-35_signal	a conserved hexamer about 35 bp upstream of the start point of bacterial transcription units; consensus=TTGACa [ ] or TGTTGACA [ ]

**Table 6: List of Feature Keys Related to Protein Sequences**

Key	Description
CONFLICT	different papers report differing sequences
VARIANT	authors report that sequence variants exist
VARSP LIC	description of sequence variants produced by alternative splicing
MUTAGEN	site which has been experimentally altered
MOD_RES	post-translational modification of a residue
ACETYLATION	N-terminal or other
AMIDATION	generally at the C-terminal of a mature active peptide
BLOCKED	undetermined N- or C-terminal blocking group
FORMYLATION	of the N-terminal methionine
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	of asparagine, aspartic acid, proline or lysine
METHYLATION	generally of lysine or arginine
PHOSPHORYLATION	of serine, threonine, tyrosine, aspartic acid or histidine
PYRROLIDONE CARBOXYLIC ACID	N-terminal glutamate which has formed an internal cyclic lactam
SULFATATION	generally of tyrosine
LIPID	covalent binding of a lipidic moiety
MYRISTATE	myristate group attached through an amide bond to the N-terminal glycine residue of the mature form of a protein or to an internal lysine residue
PALMITATE	palmitate group attached through a thioether bond to a cysteine residue or through an ester bond to a serine or threonine residue
FARNESYL	farnesyl group attached through a thioether bond to a cysteine residue
GERANYL-GERANYL	geranyl-geranyl group attached through a thioether bond to a cysteine residue
GPI-ANCHOR	glycosyl-phosphatidylinositol (GPI) group linked to the alpha-carboxyl group of the C-terminal residue of the mature form of a protein
N-ACYL DIGLYCERIDE	N-terminal cysteine of the mature form of a prokaryotic lipoprotein with an amide-linked fatty acid and a glyceryl group to which two fatty acids are linked by ester linkages
DISULFID	disulfide bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by an intra-chain disulfide bond; if the 'FROM' and 'TO' endpoints are identical, the disulfide bond is an interchain one and the description field indicates the nature of the cross-link
THIOLEST	thiolester bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by the thiolester bond
THIOETH	thioether bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by the thioether bond
CARBOHYD	glycosylation site; the nature of the carbohydrate (if known) is given in the description field
METAL	binding site for a metal ion; the description field indicates the nature of the metal
BINDING	binding site for any chemical group (co-enzyme, prosthetic group, etc.); the chemical nature of the group is given in the description field

SIGNAL	extent of a signal sequence (prepeptide)
TRANSIT	extent of a transit peptide (mitochondrial, chloroplastic, or for a microbody)
PROPEP	extent of a propeptide
CHAIN	extent of a polypeptide chain in the mature protein
PEPTIDE	extent of a released active peptide
DOMAIN	extent of a domain of interest on the sequence; the nature of that domain is given in the description field
CA_BIND	extent of a calcium-binding region
DNA_BIND	extent of a DNA-binding region
NP_BIND	extent of a nucleotide phosphate binding region; the nature of the nucleotide phosphate is indicated in the description field
TRANSMEM	extent of a transmembrane region
ZN_FING	extent of a zinc finger region
SIMILAR	extent of a similarity with another protein sequence; precise information, relative to that sequence is given in the description field
REPEAT	extent of an internal sequence repetition
HELIX	secondary structure: Helices, for example, Alpha-helix, 3(10) helix, or Pi-helix
STRAND	secondary structure: Beta-strand, for example, Hydrogen bonded beta-strand, or Residue in an isolated beta-bridge
TURN	secondary structure Turns, for example, H-bonded turn (3-turn, 4-turn or 5-turn)
ACT_SITE	amino acid(s) involved in the activity of an enzyme
SITE	any other interesting site on the sequence
INIT_MET	the sequence is known to start with an initiator methionine
NON_TER	the residue at an extremity of the sequence is not the terminal residue; if applied to position 1, this signifies that the first position is not the N-terminus of the complete molecule; if applied to the last position, it signifies that this position is not the C-terminus of the complete molecule; there is no description field for this key
NON_CONS	non consecutive residues; indicates that two residues in a sequence are not consecutive and that there are a number of unsequenced residues between them
UNSURE	uncertainties in the sequence; used to describe region(s) of a sequence for which the authors are unsure about the sequence assignment



## Appendix 3

### Specimen sequence listing

```

<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  cgcgctcctc  tcgcgcctct  ctctcgctct  cctctcgctc          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
    
```

<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

## 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
<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
<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 ; <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>- paires de bases (numéros des positions des paires de 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 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)
<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)

## 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	<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

Tableau 2 : liste des nucléotides modifiés

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-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
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

Tableau 3 : liste des acides aminés

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



Tableau 4 : liste des acides aminés modifiés ou peu connus

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
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

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 ("displacement loop") ; 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 ("diversity segment") 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 messenger 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	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 segment") 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 ("long terminal repeat") ; 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 répliquative, 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, scrRNA 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
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
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 organelle codée par le noyau; ce domaine joue un rôle dans l'importation post-traductionnelle de la protéine dans l'organelle
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 auteurs signalent qu'il existe des variants de la séquence
VARSP LIC	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
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

### 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  cggccctct  cgcgctcctc  tcgcgctct  ctctcgctct  cctctcgctc           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
    
```

<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



## Mitteilung des Europäischen Patentamts vom 2. Oktober über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen

Für internationale Anmeldungen nach dem PCT gilt ab 1. Juli 1998 der "Standard über die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Anmeldungen nach dem PCT"<sup>1</sup> (abgekürzt "PCT-Standard", siehe Regeln 5.2, 12.1 (d) und 13<sup>ter</sup> PCT sowie Abschnitt 208 und Anhang C der PCT-Verwaltungsvorschriften). Zum selben Zeitpunkt ist WIPO-Standard ST. 25 für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in Patentanmeldungen in Kraft getreten. Dieser neue WIPO-Standard, der die bisherigen WIPO-Standards ST. 23 und ST. 24 ersetzt, empfiehlt den Patentämtern, den PCT-Standard entsprechend auf alle Patentanmeldungen, bei denen es sich nicht um internationale Anmeldungen handelt, anzuwenden.

Gemäß Beschluß des Präsidenten des EPA vom 2. Oktober 1998<sup>2</sup> (nachfolgend "Beschluß des Präsidenten" genannt) ist der WIPO-Standard ST. 25 und damit der PCT-Standard für das europäische Patenterteilungsverfahren ab dem 1. Januar 1999 maßgebend.

Diese Mitteilung ergänzt den Beschluß des Präsidenten und erläutert das Verfahren nach dem EPÜ und dem PCT.

### I. Allgemeines

Seit dem 1. Januar 1993 ist in Verfahren vor dem EPA nach dem EPÜ und dem PCT die Darstellung von Nucleotid- und Aminosäuresequenzen in Form eines standardisierten Sequenzprotokolls sowie dessen Einreichung auf einem elektronischen Datenträger zwingend vorgeschrieben. Der neue WIPO Standard ST. 25 ersetzt die bisherigen Standards ST. 23 und ST. 24. Damit werden neben der Berücksichtigung des technologischen Fortschritts Vereinfachungen, Klarstellungen und Erleichterungen für die Anmelder

<sup>1</sup> Abgedruckt in der PCT-Gazette (Sonderausgabe) vom 25.6.1998.

<sup>2</sup> Abgedruckt in dieser Beilage auf Seite 1.

## Notice from the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings

The "Standard for the presentation of nucleotide and amino acid sequence listings in international applications under the PCT"<sup>1</sup> (or "PCT Standard", see Rules 5.2, 12.1(d) and 13<sup>ter</sup> PCT as well as Section 208 of and Annex C to the PCT Administrative Instructions) came into force for international applications under the PCT on 1 July 1998. WIPO Standard ST.25 for the presentation of nucleotide and amino acid sequences in patent applications also came into force on that date. This new WIPO Standard, which replaces WIPO Standards ST.23 and ST.24, recommends that patent offices also apply the PCT Standard mutatis mutandis to all patent applications in addition to international applications.

According to the decision of the President of the European Patent Office dated 2 October 1998 (hereinafter referred to as the "decision of the President"), WIPO Standard ST.25, and hence the PCT Standard, will enter into force for the European patent grant procedure on 1 January 1999.

This notice supplements the decision of the President and is intended to clarify the EPC and PCT procedures.

### I. General remarks

The presentation of nucleotide and amino acid sequences in the form of a standardised sequence listing as well as the filing thereof on an electronic data carrier has been mandatory in proceedings before the EPO under the EPC and the PCT since 1 January 1993. The new WIPO Standard ST.25 replaces Standards ST.23 and ST.24. It not only takes account of advances in technology, but also simplifies, clarifies and facilitates procedures for the applicant. In particular, it will no longer be necessary to translate a sequence

<sup>1</sup> See special edition of the PCT Gazette of 25 June 1998.

<sup>2</sup> See page 1 of this supplement.

## Communiqué de l'Office européen des brevets en date du 2 octobre 1998, relatif à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences

A compter du 1<sup>er</sup> juillet 1998, la norme applicable aux demandes internationales au titre du PCT est 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 PCT"<sup>1</sup> (en abrégé "norme du PCT", cf. règles 5.2, 12.1 d) et 13<sup>ter</sup> PCT ainsi que l'instruction administrative 208 et l'annexe C des Instructions administratives du PCT). A la même date est entrée en vigueur la norme ST. 25 de l'OMPI relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet. Cette nouvelle norme de l'OMPI, qui remplace les normes ST. 23 et ST. 24 de l'OMPI qui étaient applicables jusqu'ici, recommande aux offices de brevets d'appliquer par analogie la norme du PCT pour toutes les demandes de brevet autres que les demandes internationales.

Conformément à la décision du Président de l'OEB en date du 2 octobre 1998<sup>2</sup> (dénommée ci-après "décision du Président"), la norme ST. 25 de l'OMPI (et donc la norme du PCT) est applicable à compter du 1<sup>er</sup> janvier 1999 pour la procédure de délivrance de brevets européens.

Le présent communiqué complète la décision du Président et fournit des explications concernant la procédure selon la CBE et selon le PCT.

### I. Généralités

Dans les procédures au titre de la CBE ou du PCT se déroulant devant l'OEB, il est exigé impérativement depuis le 1<sup>er</sup> janvier 1993 la présentation d'un listage des séquences de nucléotides et d'acides aminés sous forme de listage des séquences normalisé, et le dépôt de ce listage sur un support électronique de données. La nouvelle norme ST. 25 de l'OMPI remplace les normes ST. 23 et ST. 24 qui s'appliquaient jusque là. Il est tenu compte ainsi du progrès technique, et de plus la situation se voit simplifiée, clarifiée et facilitée pour

<sup>1</sup> Reproduite dans la Gazette du PCT, numéro spécial du 25 juin 1998.

<sup>2</sup> Reproduite dans le présent supplément, page 1.

eingeführt. Insbesondere soll der Zwang zur Übersetzung eines Sequenzprotokolls bei Verwendung für verschiedene Anmeldungen in verschiedenen Sprachen dadurch entfallen, daß das Sequenzprotokoll unter Verzicht auf unnötige sprachliche Angaben (ersetzt durch numerische Codes) und unter Verwendung eines kontrollierten, von der Fachwelt allgemein anerkannten Vokabulars formuliert wird.

## II. Europäisches Patenterteilungsverfahren

### 1. Sequenzprotokoll

1.1 Die Darstellung von Nucleotid- und Aminosäuresequenzen im vorgeschriebenen Sequenzprotokoll muß dem im Anhang zum Beschluß des Präsidenten wiedergegebenen WIPO-Standard ST. 25 und damit dem PCT-Standard entsprechen (im folgenden "Standard" genannt).

1.2 Ein Sequenzprotokoll ist nur für solche Sequenzen einzureichen, die der Definition laut Nummer 2 (ii) des Standards entsprechen und sich auf die in der Anmeldung offenbarte Erfindung beziehen.

Soweit in der Beschreibung auf Sequenzen Bezug genommen wird, die zum Stand der Technik gehören und in allgemein zugänglichen Datenbanken auffindbar sind, ist es nicht erforderlich, solche Sequenzen in das Sequenzprotokoll aufzunehmen. Jedoch ist jede solche Sequenz hinreichend durch Angabe ihrer Datenbank-Zugangsnummer zu identifizieren.

1.3 Wird ein Sequenzprotokoll zusammen mit den Anmeldungsunterlagen, d. h. als Bestandteil der Beschreibung eingereicht, so sollen die Sequenzen nicht in der Beschreibung wiederholt werden (siehe Standard, Nummer 3).

1.4 Das Sequenzprotokoll kann entweder in der Sprache der Anmeldung oder (bevorzugt) in Englisch abgefaßt werden. Sind Sequenzprotokoll einerseits und Beschreibung und Ansprüche andererseits nicht in derselben Sprache abgefaßt und enthält das Sequenzprotokoll "freien Text" im Sinne des Standards (Nummer 33), so wird empfohlen, den freien Text am Ende der Beschreibung in der Sprache der Beschreibung gemäß Nummer 36 des Standards wiederzugeben (siehe Beispiele in der Anlage zu dieser Mitteilung). Damit kann vermieden werden, daß Merkmale, die nur im sogenannten "freien Text" des Sequenzprotokolls angegeben sind, in einzelnen Ver-

listing used for different applications in different languages, as sequence listings can be formulated without unnecessary text (replaced by numerical codes) and using a limited vocabulary of terms generally recognised in the field.

## II. European patent grant procedure

### 1. Sequence listing

1.1 The presentation of nucleotide and amino acid sequences in the prescribed sequence listing must comply with WIPO Standard ST.25, reproduced in the annex to the decision of the President, and thus with the PCT Standard (hereinafter referred to as the "Standard").

1.2 A sequence listing need only be filed for those sequences matching the definition given in the Standard, paragraph 2 (ii), and relating to the invention disclosed in the application.

Where the description refers to sequences which belong to the prior art and can be found in generally accessible databases, it will not be necessary to include such sequences in the sequence listing. However, each such sequence must be identified adequately by its database access number.

1.3 Where a sequence listing is filed together with the application documents, ie as part of the description, the sequences should not be repeated in the description (see Standard, paragraph 3).

1.4 The sequence listing may be written either in the language of filing or (preferably) in English. Where the sequence listing on the one hand and the description and claims on the other are not in the same language, and the sequence listing contains "free text" within the meaning of the Standard (paragraph 33), it is recommended that the free text be repeated at the end of the description in the language of the description in accordance with paragraph 36 of the Standard (see examples in the annex to this notice). This will ensure that features which are given only in the so-called "free text" of the sequence listing are not deemed in individual contracting states to the EPC not to

les demandeurs. Il n'y aura notamment plus lieu d'exiger la traduction d'un même listage des séquences figurant dans différentes demandes rédigées dans des langues différentes. Du fait que les indications linguistiques superflues disparaîtront du listage des séquences (elles seront remplacées par des codes numériques) et qu'il sera utilisé un vocabulaire contrôlé, universellement reconnu par les hommes du métier.

## II. Procédure de délivrance des brevets européens

### 1. Listage des séquences

1.1 La présentation du listage prescrit des séquences de nucléotides et d'acides aminés doit être conforme à la norme ST. 25 de l'OMPI reproduite dans l'annexe à la décision du Président, et donc à la norme du PCT (dénommée ci-après "la norme").

1.2 Il n'y a lieu de déposer de listage des séquences que pour les séquences correspondant à la définition figurant au paragraphe 2(ii) de la norme et se rapportant à l'invention exposée dans la demande.

Dans la mesure où il est fait référence dans la description à des séquences qui sont comprises dans l'état de la technique et que l'on peut trouver dans des banques de données accessibles au public, il n'est pas nécessaire d'intégrer ces séquences dans le listage des séquences. Toutefois il convient d'identifier suffisamment toutes ces séquences en indiquant leur numéro d'accès dans la banque de données.

1.3 Si un listage des séquences est déposé en même temps que les pièces de la demande, c'est-à-dire en tant que partie de la description, il n'y a pas lieu de reprendre les séquences dans la description (cf. norme, paragraphe 3).

1.4 Le listage des séquences peut être rédigé soit dans la langue de la demande, soit (de préférence) en anglais. Si le listage des séquences, d'une part, et la description et les revendications, d'autre part, ne sont pas rédigés dans la même langue, et que le listage des séquences ne comporte pas de "texte libre" au sens de la norme (paragraphe 33), il est recommandé de répéter le texte libre à la fin de la description, dans la langue de la description, conformément au paragraphe 36 de la norme, ce qui permettra d'éviter que dans certains Etats parties à la CBE des caractéristiques qui ne sont indiquées que dans le "texte libre" du listage des séquences ne soient pas

tragsstaaten des EPÜ aufgrund der Sprache nicht als Gegenstand der ursprünglichen Offenbarung angesehen werden.

## 2. Elektronischer Datenträger

2.1 Die Einreichung des Sequenzprotokolls auf einem elektronischen Datenträger zusätzlich zur schriftlichen Form ist vorgeschrieben. Die näheren Einzelheiten über die zugelassenen Datenträger, ihre Kennzeichnung, das Format der darauf gespeicherten Information und die erforderliche Hardware sind dem Standard (Nummern 39 ff.) zu entnehmen.

2.2 Das EPA empfiehlt die Verwendung der Software **Patentin**, die überarbeitet worden ist (Version 2.0) und den Anmeldern kostenlos zur Verfügung steht. Durch Verwendung dieser Software (in der jeweils gültigen Version) kann der Anmelder erreichen, daß die Formerfordernisse nach Regel 27a (1) EPÜ erfüllt sind. Die Software kann von der Website des EPA heruntergeladen werden und eignet sich für PCs, die mit dem Betriebssystem Windows (Version 3.1 oder höher) ausgestattet sind. Weitere Informationen sind erhältlich bei:

Europäisches Patentamt  
Programm STRAND  
Patentlaan 2  
NL-2288 EE Rijswijk  
Fax: (+31-70) 340 3992  
E-mail: Patentin\_help.europatent.org

### 2.3 Beschädigter oder unlesbarer Datenträger

Ist der zusammen mit der Anmeldung eingereichte Datenträger beschädigt oder aus einem anderen Grund nicht lesbar, so wird der Anmelder aufgefordert, einen einwandfreien Datenträger innerhalb einer nicht verlängerbaren Frist von einem Monat ab Zustellung der entsprechenden Mitteilung nachzureichen.

### 3. Erklärung gemäß Regel 27a (2) EPÜ

Der vorgeschriebene Datenträger wird für die EDV-gestützte europäische oder internationale Recherche sowie, nach der Veröffentlichung der Anmeldung, für die laufende Aktualisierung der internationalen Patent-Sequenzdatenbank benötigt.

Die auf dem Datenträger gespeicherte Information muß mit dem schriftlichen Sequenzprotokoll über-

be the subject-matter of the original disclosure because of the language used.

## 2. Electronic data carriers

2.1 Sequence listings must be submitted on an electronic data carrier as well as in written form. Further details with regard to authorised data carriers, their identification, the format for the information recorded on them and the necessary hardware are set out in the Standard (paragraphs 39 ff).

2.2 The EPO recommends the use of **Patentin** software, which has been upgraded (version 2.0) and is available to applicants free of charge. By using the (applicable) version of this software, applicants can ensure that the physical requirements under Rule 27a(1) EPC are met. The software can be downloaded from the EPO's website and is suitable for use with Windows 3.1 or above. Further information can be obtained from:

European Patent Office  
STRAND program  
Patentlaan 2  
NL-2288 EE Rijswijk  
Fax: (+31-70) 340 3992  
e-mail: Patentin\_help.europatent.org

### 2.3 Damaged or unreadable data carriers

If a data carrier submitted to the EPO is damaged, or unreadable for some other reason, the applicant will be invited to submit a flawless replacement within a non-extendable period of one month from notification of the communication.

### 3. Statement under Rule 27a(2) EPC

The prescribed data carrier is required for EDP-based European or international searches and, after publication of the application, for continuous updating of the international patent sequence database.

The information recorded on the data carrier must be identical to the written sequence listing. This must

considérées comme l'objet de la divulgation initiale, en raison de la différence existant entre la langue utilisée dans la description et celle utilisée dans le texte libre.

## 2. Support électronique de données

2.1 La production du listage des séquences sur un support électronique de données en plus du listage imprimé est obligatoire. Des précisions concernant les supports électroniques admis, leur identification, le format des informations y figurant et le matériel nécessaire sont données dans la norme (paragraphes 39 s.)

2.2 L'OEB recommande l'utilisation du logiciel **Patentin**, qui a été revu (version 2.0) et est mis gratuitement à la disposition des demandeurs. Le demandeur qui utilise ce logiciel (dans la version en vigueur) est en mesure de satisfaire aux conditions de forme prescrites à la règle 27bis (1) CBE. Le logiciel peut être téléchargé à partir du site web de l'OEB et convient pour les PC qui sont équipés du programme d'exploitation Windows (version 3.1 ou versions suivantes). Pour plus d'informations s'adresser à :

Office européen des brevets  
Programme STRAND  
Patentlaan 2  
NL-2288 EE Rijswijk  
Fax: (+31-70) 340 3992  
E-mail : Patentin\_help.europatent.org

### 2.3 Support de données endommagé ou illisible

Si le support de données déposé en même temps que la demande est endommagé ou illisible pour quelque autre raison, le demandeur est invité à déposer un support de données en parfait état dans un délai d'un mois à compter de la signification de la notification correspondante ; ce délai ne peut être prolongé.

### 3. Déclaration prévue à la règle 27bis (2) CBE

Le support de données prescrit est nécessaire pour la recherche européenne ou internationale assistée par ordinateur ainsi que, après la publication de la demande, pour la mise à jour constante de la banque internationale de données relatives aux séquences contenues dans les brevets.

L'information figurant sur ce support doit être identique à celle que contient le listage imprimé des séquen-

einstimmen. Der Anmelder oder sein Vertreter hat dies sicherzustellen und eine entsprechende Erklärung gemäß Regel 27a (2) EPÜ zusammen mit dem Datenträger einzureichen (siehe auch Standard, Nummer 40). Eine solche Erklärung ist im Erteilungsantrag enthalten (EPA Form 1001, Feld 31)<sup>3</sup>. Sie kann als Mustererklärung auch bei Nachreichung des Datenträgers verwendet werden.

Stellt die Prüfungsabteilung fest, daß der eingereichte Datenträger das Identitätserfordernis nicht erfüllt und somit die Erklärung gemäß Regel 27a (2) EPÜ unrichtig ist, so kann dies zur Zurückweisung der Anmeldung nach Artikel 97 (1) EPÜ führen. Darüber hinaus kann ein anderer Stand der Technik einschlägig sein, als er der Prüfung zugrunde gelegt wurde. Dies kann zu einer anderen Beurteilung der Patentierbarkeitsvoraussetzungen durch die jeweils zuständigen Instanzen auch nach Patenterteilung führen.

#### 4. Nachreichung eines Sequenzprotokolls

4.1 Ist ein Sequenzprotokoll einzureichen, so ist es den Anmeldeunterlagen als Teil der Beschreibung beizufügen (siehe Standard, Nummer 3).

4.2 Gleichwohl nimmt das EPA auch nach dem Anmeldetag eingereichte Sequenzprotokolle entgegen, sofern der vorgeschriebene Datenträger und die Erklärungen gemäß Regel 27a (2) und (3) EPÜ vorliegen, weil dies für die Durchführung der Recherche und die Fortschreibung der internationalen Patent-Sequenzdatenbank sinnvoll ist.

4.3 Ein später eingereichtes Sequenzprotokoll ist jedoch nicht Bestandteil der Beschreibung und kann daher für die Bestimmung des ursprünglichen Offenbarungsgehalts der Anmeldung nicht herangezogen werden (Regel 27a (4) EPÜ).

4.4 Ist der Anmelder nicht imstande, das Sequenzprotokoll mit der Anmeldung einzureichen, sollte er es von sich aus baldmöglichst nachreichen und die weiteren Erfordernisse nach Regel 27a (2) und (3) EPÜ erfüllen (Einreichung des vorgeschriebenen Datenträgers und der vorgesehenen Erklärungen, siehe auch Punkt II.5). Andernfalls fordert ihn die Eingangsstelle auf, die festgestellten Mängel innerhalb einer **Frist von zwei Mona-**

be ensured by the applicant or his representative, who must file together with the data carrier a statement in accordance with Rule 27a(2) EPC (see also Standard, paragraph 40). The statement is contained in the Request for Grant (EPO Form 1001, section 31)<sup>3</sup>, and may also be used as a model statement when the data carrier is filed subsequently.

If the examining division finds that the data carrier submitted does not comply with the requirement that it be identical and the statement under Rule 27a (2) EPC is therefore incorrect, this may lead to refusal of the application under Article 97(1) EPC. Furthermore, a different prior art than that on which the examination was based may be relevant. This may lead to a different assessment by the relevant departments or courts of the conditions for patentability, even after a patent has been granted.

#### 4. Subsequent filing of a sequence listing

4.1 If a sequence listing has to be filed, it should be annexed to the application documents as part of the description (see Standard, paragraph 3).

4.2 Nevertheless, the EPO accepts sequence listings filed after the date of filing so long as the prescribed data carrier and the statements under Rule 27a(2) and (3) EPC have been received because they can be useful in the search and for updating the international patent sequence database.

4.3 A sequence listing filed later does not, however, form part of the description and cannot therefore be used in determining the original disclosed content of the application (Rule 27a(4) EPC).

4.4 If the applicant is unable to file the sequence listing with the description, he should file it without prompting as soon as possible thereafter and comply with the other requirements under Rule 27a(2) and (3) EPC (see also point II.5 for submission of the prescribed data carrier and statements). Otherwise, the Receiving Section will invite him to remedy the deficiencies within a **period of two months** (Article 91(2),

ces. Le demandeur ou son mandataire est tenu de prendre les mesures nécessaires à cet effet et de joindre au support de données une déclaration en ce sens, conformément à la règle 27bis (2) CBE (cf. également la norme, paragraphe 40). Une telle déclaration figure dans la requête en délivrance (OEB Form 1001, rubrique 31)<sup>3</sup>. Elle peut également être utilisée comme déclaration type en cas de production ultérieure du support de données.

Si la division d'examen constate que le support de données ne satisfait pas à l'exigence d'identité avec la liste écrite et que la déclaration selon la règle 27bis (2) CBE est donc inexacte, la demande peut être rejetée conformément à l'article 97(1) CBE. Il se peut en outre qu'un état de la technique autre que celui qui a été pris en compte pour l'examen soit pertinent et qu'il en résulte une appréciation divergente par les instances compétentes des conditions de brevetabilité même après la délivrance du brevet.

#### 4. Dépôt ultérieur d'un listage des séquences

4.1 Tout listage des séquences doit être joint aux pièces de la demande comme faisant partie de la description (cf. norme, paragraphe 3).

4.2 L'OEB accepte néanmoins les listages déposés après la date de dépôt, dans la mesure où le support de données prescrit et les déclarations selon la règle 27bis (2) et (3) CBE ont été produits, étant donné que ces informations sont utiles pour effectuer la recherche et actualiser la banque internationale de données relatives aux séquences contenues dans les brevets.

4.3 Un listage de séquences déposé ultérieurement ne fait toutefois pas partie de la description et ne peut donc servir à déterminer le contenu initial de la demande (règle 27bis (4) CBE).

4.4 Si le demandeur n'est pas en mesure de déposer le listage des séquences avec la demande, il a intérêt à en effectuer spontanément le dépôt ultérieurement dans les meilleurs délais, en veillant à remplir les autres conditions énoncées à la règle 27bis (2) et (3) CBE (production du support de données prescrit et des déclarations prévues; cf. également point II.5). Sinon, la section de dépôt l'invite à remédier dans un

<sup>3</sup> Abgedruckt im ABI. EPA 1997, 251.

<sup>3</sup> Published in OJ EPO 1997, 251.

<sup>3</sup> Reproduit au JO OEB 1997, 251.

ten zu beheben (Art. 91 (2), Regel 41 (1) EPÜ und Artikel 3 des Beschlusses des Präsidenten). Die Frist wird **einmal um höchstens zwei Monate** verlängert, sofern der Anmelder dies vor Ablauf der Frist beantragt. In begründeten Ausnahmefällen kann die Frist auch von Amts wegen verlängert werden, wenn nicht alle Mängel innerhalb der gesetzten Frist beseitigt worden sind. Wird innerhalb dieser Frist ein **beschädigter oder aus einem anderen Grund unlesbarer Datenträger** eingereicht, so wird dem Anmelder eine besondere, nicht verlängerbare Frist von einem Monat ab Zustellung der entsprechenden Mitteilung zur Nachreichung eines einwandfreien Datenträgers eingeräumt.

##### 5. Erklärung gemäß Regel 27a (3) EPÜ

Ein nachgereichtes Sequenzprotokoll darf nichts enthalten, das über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht, da die Recherche auf der Grundlage des nachgereichten Sequenzprotokolls durchgeführt wird. Das nachgereichte Sequenzprotokoll darf daher nur diejenige Sequenzinformation – in standardisierter Form – enthalten, die in den ursprünglichen Anmeldungsunterlagen bereits vorhanden war. Es ist unbedingt darauf zu achten, daß das nachgereichte Sequenzprotokoll nicht mehr und nicht weniger Sequenzen als die in der ursprünglichen Beschreibung offenbarten Sequenzen enthält und die Numerierung dieser Sequenzen unverändert bleibt. Der Anmelder hat dies sicherzustellen und eine entsprechende Erklärung gemäß Regel 27a (3) EPÜ einzureichen. Es wird empfohlen, hierfür die in der Fußnote abgedruckte Mustererklärung<sup>4</sup> zu verwenden. Für den ebenfalls einzureichenden Datenträger gilt Punkt II.3.

##### 6. Rechtsverlust

Werden die Erfordernisse der Regel 27a EPÜ in Verbindung mit dem Beschluß des Präsidenten nicht rechtzeitig, gegebenenfalls nach Aufforderung der Eingangsstelle gemäß Regel 41 (1) EPÜ, erfüllt, so wird die Anmeldung zurückgewiesen (Art. 91 (1) b) und (3) EPÜ).

Der Anmelder kann die Weiterbehandlung der Anmeldung nach Artikel 121 EPÜ beantragen.

Rule 41(1) EPC and Article 3 of the decision of the President). This period may be extended **once only for a maximum of two months**, provided that the applicant requests such an extension before the period expires. In exceptional, substantiated cases, the period may be extended by the Receiving Section of its own motion where not all the deficiencies have been remedied within the prescribed period. If a **data carrier is filed which is damaged, or unreadable for some other reason**, the applicant will be granted a special, non-extendable period of one month from notification of the communication within which to file a flawless replacement.

##### 5. Statement under Rule 27a(3) EPC

A subsequently filed sequence listing may not include any matter which goes beyond the content of the application as filed, since the search is to be carried out on the basis of the subsequently filed sequence listing. The subsequently filed sequence listing may therefore contain only the sequence information – in a standardised form – that was already in the original application documents. The subsequently filed sequence listing should contain no more and no fewer sequences than disclosed in the original description, and their numbering should remain unchanged. This must be ensured by the applicant, who must file a statement to that effect in accordance with Rule 27a(3) EPC. It is recommended that the specimen statement in footnote<sup>4</sup> below be used for this purpose. Point II.3 is applicable for the data carrier which must also be submitted.

##### 6. Loss of rights

If the requirements laid down in Rule 27a EPC in conjunction with the decision of the President are not met in good time – if applicable, after an invitation from the Receiving Section under Rule 41(1) EPC – the application is refused (Art. 91(1)(b) and (3) EPC).

The applicant may request further processing of the application under Article 121 EPC.

**délai de deux mois** aux irrégularités constatées (art. 91(2), règle 41(1) CBE et art. 3 de la décision du Président). Le délai est prolongé **une seule fois de deux mois au maximum**, si le demandeur présente une requête en ce sens avant l'expiration du délai. Dans des cas exceptionnels, si la situation le justifie, le délai peut également être prolongé d'office, s'il n'a pas été remédié à toutes les irrégularités dans le délai fixé. S'il est déposé dans ce délai un **support de données endommagé ou rendu illisible pour quelque autre raison**, il est imparti au demandeur un délai spécial non prorogeable d'un mois à compter de la signification de la notification correspondante invitant le demandeur à déposer un support de données en parfait état.

##### 5. Déclaration prévue à la règle 27bis (3) CBE

Un listage des séquences déposé ultérieurement ne doit contenir aucun élément s'étendant au-delà du contenu de la demande telle qu'elle a été déposée, puisque la recherche est exécutée sur la base de ce listage. Le listage des séquences déposé ultérieurement ne doit contenir – sous forme normalisée – que les informations concernant les séquences qui figuraient déjà dans les pièces initiales de la demande. Il faut absolument veiller à ce que le listage des séquences déposé ultérieurement ne contienne ni plus ni moins de séquences que celles qui sont divulguées dans la description initiale et à ce que la numérotation de ces séquences demeure inchangée. Le demandeur est tenu de faire le nécessaire à cet effet et de produire une déclaration en ce sens conformément à la règle 27bis (3) CBE. Il est recommandé d'utiliser à cette fin la déclaration type reproduite dans la note en bas de page<sup>4</sup>. Le point II.3 est applicable au support de données que le demandeur doit en outre produire.

##### 6. Perte de droits

Si le demandeur ne satisfait pas en temps utile aux conditions énoncées à la règle 27bis CBE en liaison avec la décision du Président de l'OEB, le cas échéant après y avoir été invité par la section de dépôt conformément à la règle 41 (1) CBE, la demande est rejetée (art. 91(1)(b) et (3) CBE).

Le demandeur peut demander la poursuite de la procédure conformément à l'article 121 CBE.

<sup>4</sup> "Es wird hiermit erklärt, daß das Sequenzprotokoll nicht über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht."

<sup>4</sup> "It is hereby declared that the sequence listing does not go beyond the content of the application as filed."

<sup>4</sup> "Il est déclaré par la présente que le listage des séquences ne s'étend pas au-delà du contenu de la demande telle qu'elle a été déposée."

### III. Verfahren vor dem EPA als Internationaler Recherchenbehörde oder als mit der internationalen vorläufigen Prüfung beauftragter Behörde

1. Sequenzprotokolle in maschinenlesbarer Form werden für die Durchführung der internationalen Recherche ebenso benötigt wie für die europäische Recherche. Sie müssen dem PCT-Standard entsprechen (Regel 5.2 PCT sowie Abschnitt 208 und Anhang C der PCT-Verwaltungsvorschriften).

Liegt ein diesem Standard entsprechendes Sequenzprotokoll zu Beginn der internationalen Recherche nicht oder nicht auf elektronischem Datenträger vor, so wird der Anmelder vom EPA aufgefordert, das Sequenzprotokoll und/oder den vorgeschriebenen Datenträger innerhalb einer **nicht verlängerbaren Frist von maximal zwei Monaten** nachzureichen (Regel 13<sup>ter</sup>.1 a) PCT). Diese wird wegen der engen Fristen der Regel 42 PCT für die Erstellung des Recherchenberichts regelmäßig nur **einen Monat** betragen.

2. Bei Nachreichung des Sequenzprotokolls und/oder des Datenträgers hat der Anmelder zu erklären, daß das nachgereichte Sequenzprotokoll nicht über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht und/oder die auf dem Datenträger gespeicherte Information mit dem schriftlichen Sequenzprotokoll übereinstimmt (Art. 4 des Beschlusses des Präsidenten und Standard, Nummern 37 und 40).

3. Werden diese Erfordernisse nicht oder nicht rechtzeitig erfüllt, so führt das EPA keine internationale Recherche durch, soweit eine sinnvolle Recherche nicht möglich ist (Regel 13<sup>ter</sup>.1 c) PCT). In diesem Fall wird der internationale Recherchenbericht ganz oder teilweise durch die Erklärung gemäß Artikel 17 (2) PCT ersetzt.

4. Ist der internationale Recherchenbericht aus den vorgenannten Gründen nicht oder nur teilweise erstellt worden, so führt das EPA als mit der internationalen vorläufigen Prüfung beauftragte Behörde keine internationale vorläufige Prüfung durch, soweit eine sinnvolle Prüfung nicht möglich ist (Regel 66.1 e) PCT). Dies wird dem Anmelder mitgeteilt und im internationalen vorläufigen Prüfungsbericht vermerkt (Regel 66.2 a) vii)

### III. Proceedings before the EPO as International Searching Authority or International Preliminary Examining Authority

1. Sequence listings in machine readable form are just as necessary in carrying out the international search as they are for the European search. They must comply with the PCT Standard (Rule 5.2 PCT as well as Section 208 of and Annex C to the PCT Administrative Instructions).

If at the start of the international search a sequence listing complying with this Standard has not been received or has not been received on an electronic data carrier, the EPO invites the applicant to file the sequence listing and/or the prescribed data carrier within a **non-extendable period of not more than two months** (Rule 13<sup>ter</sup>.1(a) PCT). Because of the short time limits in Rule 42 PCT for the establishment of the search report, this period will normally be only **one month**.

2. If the sequence listing and/or data carrier is filed subsequently, the applicant must state that the sequence listing so filed does not include matter which goes beyond the content of the application as filed and/or that the information recorded on the data carrier is identical to the written sequence listing (Article 4 of the decision of the President, and Standard, paragraphs 37 and 40).

3. If these requirements are not met or not met in due time, the EPO does not perform the international search where a meaningful search cannot be carried out (Rule 13<sup>ter</sup>.1(c) PCT). In this case the international search report is replaced in full or in part by the statement under Article 17(2) PCT.

4. If the international search report has not been drawn up in full or in part for the aforementioned reasons, the EPO as International Preliminary Examining Authority does not perform the international preliminary examination where a meaningful examination cannot be carried out (Rule 66.1(e) PCT). This is notified to the applicant and stated in the international preliminary examination report (Rules 66.2(a)(vii) and 70.12(iv))

### III. Procédure devant l'OEB agissant en qualité d'administration chargée de la recherche internationale ou d'administration chargée de l'examen préliminaire international

1. Les listages des séquences sous forme déchiffirable par machine sont nécessaires tant pour la recherche internationale que pour la recherche européenne. Ils doivent être établis conformément à la norme du PCT (règle 5.2 PCT ainsi que instruction administrative 208 et annexe C des Instructions administratives du PCT).

Si un listage des séquences conforme à cette norme n'est pas parvenu à l'OEB ou bien s'il n'est pas disponible sur support électronique de données lorsque la recherche internationale est entreprise, le demandeur est invité par l'OEB à fournir **dans un délai non prorogable de deux mois au maximum** le listage et/ou le support de données prescrit (règle 13<sup>ter</sup>.1 a) PCT). Ce délai ne sera normalement que d'**un mois** en raison des délais très stricts fixés par la règle 42 PCT pour l'établissement du rapport de recherche internationale.

2. Lorsqu'il fournit ultérieurement le listage des séquences et/ou le support de données, le demandeur est tenu de déclarer que ce listage ne s'étend pas au-delà du contenu de la demande telle qu'elle a été déposée et/ou que l'information figurant sur ce support est identique à celle que contient le listage imprimé (art. 4 de la décision du Président et norme, paragraphes 37 et 40).

3. Si ces conditions ne sont pas remplies ou ne le sont pas dans les délais, l'OEB ne procède pas à la recherche internationale dans la mesure où une recherche significative ne peut pas être effectuée (règle 13<sup>ter</sup>.1 c) PCT). Dans ce cas, le rapport de recherche internationale est remplacé en tout ou en partie par la déclaration prévue à l'article 17(2) PCT.

4. Si le rapport de recherche internationale n'a pas été établi, ou n'a été établi qu'en partie, pour les raisons qui viennent d'être énoncées, l'OEB agissant en qualité d'administration chargée de l'examen préliminaire international ne procède pas à l'examen préliminaire international dans la mesure où un examen significatif ne peut pas être effectué (règle 66.1 e) PCT). L'OEB en informe le demandeur et l'indique dans le rapport

und 70.12 iv) PCT). Eine Aufforderung zur Einreichung eines Sequenzprotokolls gemäß Regel 13<sup>ter</sup>.1 (e) PCT ergeht nicht.

Es wird nachdrücklich davon abgeraten, ein Sequenzprotokoll erst in diesem späten Stadium einzureichen. Zumindest bei umfangreichen Sequenzen wäre das EPA nicht in der Lage zu überprüfen, ob das Sequenzprotokoll über den Offenbarungsgehalt der internationalen Anmeldung im Anmeldezeitpunkt hinausgeht. Zu einer derartigen Überprüfung ist es nach Artikel 34 (2) b) und Regel 66.2 (a) iv) PCT zwar verpflichtet. Anders als bei der internationalen Recherche kann eine Erklärung des Anmelders, daß das nachgereichte Sequenzprotokoll nicht über den Offenbarungsgehalt der internationalen Anmeldung hinausgeht, das EPA nicht von einer eigenen Überprüfung dieses Tatbestandes (Artikel 34 (2) b), Regel 66.2 a) PCT) entbinden. Es gibt aber derzeit keine geeigneten Mittel, die einen zufriedenstellenden Vergleich der ursprünglichen Anmeldungsunterlagen mit dem nachgereichten Sequenzprotokoll zu vertretbarem Aufwand ermöglichen könnten. Daher könnte der internationale vorläufige Prüfungsbericht nicht auf der Grundlage des nachgereichten Sequenzprotokolls, sondern nur auf der Grundlage der am Anmeldetag in der Beschreibung vorhandenen Sequenzen erstellt werden.

#### IV. Eintritt der Euro-PCT-Anmeldungen in die regionale Phase vor dem EPA

1. Liegt ein nach Regel 5.2 PCT vorgeschriebenes Sequenzprotokoll bei Ablauf der internationalen Phase dem EPA als Bestimmungsamt oder ausgewähltem Amt nicht vor, entspricht es nicht dem Standard oder ist der vorgeschriebene Datenträger (siehe Standard, Nummern 39 ff.) nicht eingereicht worden, so wird der Anmelder aufgefordert, die festgestellten Mängel innerhalb einer **Frist von zwei Monaten** zu beheben (Regel 13<sup>ter</sup>.2 PCT, Regel 104b (3a) EPÜ und Artikel 5 des Beschlusses des Präsidenten). Gegebenenfalls wird der Anmelder aufgefordert, gleichzeitig zu erklären, daß das nachgereichte Sequenzprotokoll nicht über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht und/oder die auf dem

PCT). An invitation under Rule 13<sup>ter</sup>.1(e) PCT to file a sequence listing is not issued.

Applicants are strongly advised not to file sequence listings at this late stage. In the case of long sequence listings at least, the European Patent Office would not be in a position to examine whether the sequence listing goes beyond the content of the international application as disclosed at the time of filing, although under Article 34(2)(b) and Rule 66.2(a)(iv) PCT it is obliged to carry out such an examination. Unlike the situation pertaining to the international search, a statement from the applicant to the effect that the subsequently filed sequence listing does not go beyond the content of the international application as filed cannot release the EPO from its obligation to carry out its own examination (Article 34(2)(b), Rule 66.2(a) PCT). However, there is at present no appropriate means of carrying out a satisfactory comparison of the documents as originally filed with the subsequently filed sequence listing. The international preliminary examination report therefore could not be established on the basis of the subsequently filed sequence listing but only on the basis of the sequences available in the description on the date of filing.

#### IV. Entry of Euro-PCT applications into the regional phase before the EPO

1. If on expiry of the international phase the sequence listing prescribed in Rule 5.2 PCT is not available to the EPO as designated Office or elected Office or does not comply with the Standard, or if the prescribed data carrier (see Standard, paragraphs 39 ff) has not been filed, the applicant is invited to remedy such deficiencies within a **period of two months** (Rule 13<sup>ter</sup>.2 PCT, Rule 104b(3a) EPC and Article 5 of the decision of the President). Where appropriate, the applicant is invited at the same time to state that the subsequently filed sequence listing does not go beyond the content of the application as filed and/or that the information recorded on the data carrier is identical to the written sequence listing (Standard, para-

d'examen préliminaire international (règles 66.2 a) vii) et 70.12 iv) PCT). Il n'est pas envoyé d'invitation à déposer un listage des séquences conformément à la règle 13<sup>ter</sup>.1 e) PCT.

Il est tout à fait déconseillé de ne déposer un listage des séquences qu'à ce stade tardif de la procédure. Tout au moins dans le cas de séquences longues, l'OEB ne serait pas en mesure de vérifier si le listage des séquences ne s'étend pas au-delà de l'exposé de l'invention figurant dans la demande internationale à la date du dépôt. Il est certes tenu de procéder à une telle vérification en vertu de l'article 34 (2) b) et de la règle 66.2 a) iv) PCT, mais contrairement à ce qui se passe dans le cas de la recherche internationale, il ne suffit pas que le demandeur déclare que le listage des séquences déposé ultérieurement ne s'étend pas au-delà de l'exposé de l'invention figurant dans la demande internationale à la date du dépôt pour que l'OEB puisse se dispenser de procéder lui-même à cette vérification (article 34 (2) b), règle 66.2 a) PCT). Toutefois il n'existe pas actuellement de moyen approprié permettant d'établir une comparaison satisfaisante, ne nécessitant pas une somme de travail excessive, entre les pièces initiales de la demande et le listage des séquences déposé ultérieurement. C'est pourquoi le rapport d'examen préliminaire international ne saurait être établi sur la base du listage des séquences déposé ultérieurement, mais uniquement sur la base des séquences figurant dans la description à la date du dépôt.

#### IV. Entrée des demandes euro-PCT dans la phase régionale devant l'OEB

1. Si, au terme de la phase internationale, le listage des séquences visé à la règle 5.2 PCT n'est pas parvenu à l'OEB agissant en qualité d'Office désigné ou d'Office élu, n'a pas été établi conformément à la norme ou n'a pas été produit sur le support de données prescrit (cf. norme, paragraphes 39 s.), le demandeur est invité à remédier dans un **délai de deux mois** aux irrégularités constatées (règle 13<sup>ter</sup>.2 PCT, règle 104<sup>ter</sup> (3bis) CBE et article 5 de la décision du Président). Le cas échéant, le demandeur est invité à déclarer en même temps que le listage des séquences déposé ultérieurement ne s'étend pas au-delà du contenu de la demande telle qu'elle a été déposée et/ou que l'information figurant sur le support de données est identique à

Datenträger gespeicherte Information mit dem schriftlichen Sequenzprotokoll übereinstimmt (Standard, Nummern 37 und 40). Artikel 3 (2) und (3) des Beschlusses des Präsidenten gelten entsprechend (vgl. oben Punkt II 4.4).

2. Werden bestehende Mängel nach einer solchen Aufforderung nicht oder nicht rechtzeitig beseitigt, so wird die Anmeldung zurückgewiesen (Art. 91 (3) EPÜ).

3. Der Anmelder kann die Weiterbehandlung der Anmeldung nach Artikel 121 EPÜ beantragen.

4. Übersetzung gemäß Artikel 158 (2) EPÜ

Einer Übersetzung der Textstellen eines Sequenzprotokolls, das Bestandteil der Beschreibung ist, bedarf es grundsätzlich nicht, sofern das Sequenzprotokoll der Regel 12.1 (d) PCT und die Beschreibung der Regel 5.2 (b) PCT entspricht (Regel 49.5 (a-bis) PCT). Ist die Voraussetzung nach Regel 5.2 (b) PCT nicht erfüllt, so verlangt das EPA keine solche Übersetzung, wenn das Sequenzprotokoll in englisch abgefaßt ist und die internationale Anmeldung in einer anderen EPA-Amtssprache (deutsch oder französisch) veröffentlicht worden ist.

#### V. Veröffentlichung von Sequenzprotokollen durch das EPA – Akteneinsicht

1. Ist das Sequenzprotokoll **zusammen** mit den Unterlagen der europäischen Patentanmeldung (oder internationalen Anmeldung) als Teil der Beschreibung eingereicht worden, so wird es zusammen mit der Anmeldung durch das EPA (oder durch das Internationale Büro) veröffentlicht.

2. Wird ein Sequenzprotokoll zu einer europäischen oder zu einer in der regionalen Phase vor dem EPA anhängigen internationalen Anmeldung **nachgereicht**, so hängt seine Veröffentlichung von dem Zeitpunkt der Einreichung ab (Artikel 6 des Beschlusses des Präsidenten):

Wird es rechtzeitig vor Abschluß der technischen Vorbereitungen für die Veröffentlichung der europäischen Patentanmeldung<sup>5</sup> eingereicht, so wird es als Anlage zu den Anmeldeunterlagen (A-Schrift) oder im Falle einer Euro-PCT-Anmeldung, die nicht in einer Amtssprache des EPA

graphs 37 and 40). Article 3(2) and (3) of the decision of the President applies *mutatis mutandis* (see point II 4.4 above).

2. If existing deficiencies are not remedied or not remedied in due time after such an invitation, the application is refused (Art. 91(3) EPC).

3. The applicant may request further processing of the application under Article 121 EPC.

4. Translation in accordance with Article 158(2) EPC

There is no requirement to translate any text matter in a sequence listing which forms part of the description, provided that the sequence listing complies with Rule 12.1(d) PCT and the description with Rule 5.2(b) PCT (Rule 49.5(a-bis) PCT). If the requirement under Rule 5.2(b) PCT is not met, the EPO does not require such a translation where the sequence listing is in English and the international application has been published in one of the other official languages of the EPO (French or German).

#### V. Publication of sequence listings by the EPO – inspection of files

1. If the sequence listing has been filed **together** with the documents making up the European (or international) application as part of the description, it is published together with the application by the EPO (or the International Bureau).

2. If a sequence listing pertaining to a European patent application or to an international application in the regional phase before the EPO is **subsequently filed**, its publication depends on the date of filing (Article 6 of the decision of the President):

If it is filed in due time before completion of the technical preparations for publication of the European patent application<sup>5</sup>, it is published as an annex to the application documents (A document) or, in the case of a Euro-PCT application which has not been published internationally in one

celle que contient le listage imprimé (la norme, paragraphes 37 et 40). L'article 3(2) et (3) de la décision du Président est applicable par analogie (cf. supra, point II 4.4).

2. Si, après une telle invitation, le demandeur ne remédie pas aux irrégularités constatées ou n'y remédie pas en temps utile, la demande est rejetée (art. 91(3) CBE).

3. Le demandeur peut demander la poursuite de la procédure conformément à l'article 121 CBE.

4. Traduction conformément à l'article 158(2) CBE

Il n'est pas nécessaire en tout état de cause de déposer une traduction des éléments de texte d'un listage des séquences qui fait partie de la description, si ce listage des séquences est conforme à la règle 12.1 d) PCT et si la description est conforme à la règle 5.2 b) PCT (règle 49.5 (a-bis) PCT). Si la condition requise à la règle 5.2 b) PCT n'est pas remplie, l'OEB n'exige pas une telle traduction si le listage des séquences est rédigé en anglais et si la demande internationale a été publiée dans une autre langue officielle de l'OEB (allemand ou français).

#### V. Publication par l'OEB des listages des séquences – inspection publique

1. Si le listage des séquences a été déposé **en même temps** que les pièces de la demande européenne ou internationale comme partie de la description, il est publié avec la demande par l'OEB (ou par le Bureau international).

2. Si le listage des séquences d'une demande européenne ou d'une demande internationale se trouvant en instance devant l'OEB dans la phase régionale a été **déposé ultérieurement**, sa publication dépend de la date de son dépôt (article 6 de la décision du Président).

S'il est déposé en temps utile avant la fin des préparatifs techniques entrepris en vue de la publication de la demande de brevet européen<sup>5</sup>, il est publié en annexe aux pièces de la demande (document A) ou, s'il s'agit d'une demande euro-PCT qui n'a pas fait l'objet d'une publication interna-

<sup>5</sup> 7 Wochen vor Ablauf von 18 Monaten nach dem Anmelde- oder Prioritätstag; Beschluß des Präsidenten des EPA vom 14. Dezember 1992, ABI. EPA 1993, 55.

<sup>5</sup> Seven weeks before expiry of the eighteenth month following the date of filing or priority; Decision of the President of the EPO dated 14 December 1992, OJ EPO 1993, 55.

<sup>5</sup> Sept semaines avant l'expiration d'un délai de dix-huit mois à compter de la date de dépôt ou de priorité ; décision du Président de l'OEB en date du 14 décembre 1992, JO OEB 1993, 55.



international veröffentlicht worden ist, als Anlage zu der nach Artikel 158 (3) EPÜ veröffentlichten Übersetzung mitveröffentlicht.

In allen anderen Fällen besteht nur die Möglichkeit, bei Patenterteilung das Sequenzprotokoll zusammen mit der Patentschrift zu veröffentlichen. Die Veröffentlichung des nachgereichten Sequenzprotokolls zusammen mit dem gesonderten Recherchenbericht (sogenannte A3-Schrift) ist nicht vorgesehen.

3. Ein Sequenzprotokoll, das nicht zusammen mit den Anmeldeunterlagen oder mit der Patentschrift veröffentlicht wird, ist gegebenenfalls nur durch Akteneinsicht oder über die Patent-Sequenzdatenbanken der Öffentlichkeit zugänglich. Auf Antrag wird Akteneinsicht auch durch Herstellung und Zusendung einer Kopie des vom Anmelder eingereichten elektronischen Datenträgers gewährt (siehe II.2 und IV.1 und Regel 94 (3) EPÜ).

of the official languages of the EPO, as an annex to the translation published under Article 158(3) EPC.

In all other cases it is only possible to publish the sequence listing together with the patent specification when the patent is granted. There is no provision for publishing the subsequently filed sequence listing together with the separate search report (A3 document).

3. A sequence listing which is not published together with the application documents or with the patent specification is, where applicable, accessible to the public only by way of a file inspection or by way of the patent sequence databases. On request, inspection of files may also be granted by sending the requester a copy of the electronic data carrier filed by the applicant (see II.2 and IV.1 and Rule 94(3) EPC).

tionale dans une langue officielle de l'OEB, en annexe à la traduction publiée conformément à l'article 158(3) CBE.

Dans tous les autres cas, la seule possibilité consiste, si la demande donne lieu à la délivrance d'un brevet, à publier le listage des séquences avec le fascicule de brevet. Il n'est pas prévu qu'un listage des séquences déposé ultérieurement soit publié avec le rapport de recherche séparé (document A3).

3. Un listage des séquences qui n'est publié ni avec les pièces de la demande ni avec le fascicule de brevet n'est le cas échéant accessible au public que dans le cadre de l'inspection publique ou par l'intermédiaire de la banque de données relatives aux séquences contenues dans les brevets. Il est également procédé, sur requête, à l'inspection publique, par établissement et envoi d'une copie du support de données électronique produit par le demandeur (cf. II.2 et IV.1 et règle 94(3) CBE).

**Anlage zu der Mitteilung des Europäischen Patentamts vom 2. Oktober 1998 über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen\***

**Appendix to the Notice from the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings\***

**Annexe au Communiqué de l'Office européen des brevets en date du 2 octobre 1998, relatif à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences\***

**Freier Text des Sequenzprotokolls**

**Sequence Listing Free Text**

**Texte libre du listage des séquences**

**Beispiel 1:**

**Example 1:**

**Exemple 1 :**

Dies ist ein Beispiel des Abschnitts "Freier Text des Sequenzprotokolls" (vgl. Nummer 36 des WIPO-Standards ST. 25) in der Beschreibung einer in **Französisch** eingereichten europäischen oder internationalen Patentanmeldung, die ein Sequenzprotokoll mit Freitextelementen in Englisch zur Sequenz Nr. 3 enthält:

This is an example of a "Sequence Listing Free Text" section (see paragraph 36 of WIPO Standard ST. 25) in the description of a European patent application or international application filed in **French** and containing a sequence listing with English free text data elements for Sequence No. 3:

Ceci est un exemple de la rubrique particulière de la description intitulée "texte libre du listage des séquences" (voir le paragraphe 36 de la norme OMPI ST.25) pour une demande de brevet européen ou une demande internationale déposée en **français** et contenant un listage des séquences comportant des éléments de texte libre en anglais pour la séquence n° 3 :

```
<210> 3
<223> /note="PCV"
        /hôte spécifique="Sus scrofa"
        /type de tissu="rein"
        /lignée cellulaire="PK-15"
<223> /note="TAGTATTAC homologue au circovirus de plante et au motif nonanucléotide de géminivirus"
<223> /Codon de départ =1
        /note="similaire au circovirus BBTV de plante"
        /produit="site de réplication de la protéine associée au PCV"
<223> /note="fin de réplication de la protéine associée au PCV a lieu 6 nucléotides en aval du signal
de polyadénylation
<223> /note="TAGTATTAC homologue au circovirus de plante et au motif nonanucléotide de géminivirus"
```

**Beispiel 2:**

**Example 2**

**Exemple 2 :**

Dies ist ein Beispiel des Abschnitts "Freier Text des Sequenzprotokolls" (vgl. Nummer 36 des WIPO-Standards ST. 25) in der Beschreibung einer in **Deutsch** eingereichten europäischen oder internationalen Patentanmeldung, die ein Sequenzprotokoll mit Freitextelementen in Englisch zur Sequenz Nr. 3 enthält:

This is an example of a "Sequence Listing Free Text" section (see paragraph 36 of WIPO Standard ST. 25) in the description of a European patent application or international application filed in **German** and containing a sequence listing with English free text data elements for Sequence No. 3:

Ceci est un exemple de la rubrique particulière de la description intitulée "texte libre du listage des séquences" (voir le paragraphe 36 de la norme OMPI ST.25) pour une demande de brevet européen ou une demande internationale déposée en **allemand** et contenant un listage des séquences comportant des éléments de texte libre en anglais pour la séquence n° 3 :

```
<210> 3
<223> /Anmerkung="PCV"
        /Spezifischer Wirt="Sus scrofa"
        /Gewebetyp="Niere"
        /Zelllinie="PK-15"
<223> /Anmerkung="TAGTATTAC-Sequenz, homolog zu pflanzlichem Circovirus- und Geminivirus-spezifischem
Nona-Nucleotidmotiv"
<223> /Startcodon = 1
        /Anmerkung="ähnlich dem pflanzlichen Circovirus BBTV"
        /Produkt="Replikationsstelle des assoziierten Proteins PCV"
<223> /Anmerkung="Ende des mutmaßlichen replikations-assozierten Proteins PCV tritt 6 Nucleotide stromabwärts
des potentiellen Polyadenylierungssignals auf"
<223> /Anmerkung="TAGTATTAC-Sequenz, homolog zu pflanzlichem Circovirus- und Geminivirus-spezifischem
Nona-Nucleotidmotiv"
```

\* Siehe II 1.4 der Mitteilung in dieser Beilage, S. 60.

\* See II 1.4 of the Notice in this supplement, p. 60.

\* Voir le point II 1.4 du communiqué reproduit dans le présent supplément, p. 60.





**Beschluß des Präsidenten des Europäischen Patentamts vom 2. Oktober 1998 über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen**

Der Präsident des Europäischen Patentamts, gestützt auf Artikel 10 EPÜ, Regeln 27a und 49 EPÜ sowie Regel 13<sup>ter</sup> PCT, beschließt:

**Artikel 1**

*Standardisierte Darstellung von Nucleotid- und Aminosäuresequenzen in europäischen Patentanmeldungen*

Hat die in einer europäischen Patentanmeldung offenbarte Erfindung Nucleotid- oder Aminosäuresequenzen zum Gegenstand, so hat die Beschreibung ein Sequenzprotokoll zu enthalten, das dem im **Anhang** zu diesem Beschluß wiedergegebenen WIPO-Standard ST. 25 entspricht.

**Artikel 2**

*Einreichung von Sequenzprotokollen auf elektronischem Datenträger*

(1) Zusätzlich ist ein Datenträger mit dem Sequenzprotokoll einzureichen, der dem im **Anhang** zu diesem Beschluß wiedergegebenen Standard entspricht.

(2) Wird ein Sequenzprotokoll nach seiner Einreichung beim Europäischen Patentamt berichtigt, so ist ein neuer Datenträger mit dem berichtigten Sequenzprotokoll einzureichen.

(3) Jedem Datenträger ist eine Erklärung des Anmelders beizufügen, daß die auf dem Datenträger gespeicherte Information mit dem schriftlichen Sequenzprotokoll übereinstimmt.

**Artikel 3**

*Beseitigung von Mängeln*

(1) Wird den in Regel 27a EPÜ und in den Artikeln 1 und 2 dieses Beschlusses genannten Erfordernissen nicht entsprochen, so teilt die Eingangsstelle dies dem Anmelder mit und fordert ihn auf, die festgestellten Mängel innerhalb einer Frist von zwei Monaten zu beseitigen (Regel 41 (1) EPÜ).

**Decision of the President of the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings**

The President of the European Patent Office, having regard to Article 10 EPC, Rules 27a and 49 EPC and Rule 13<sup>ter</sup> PCT, has decided as follows:

**Article 1**

*Standardised presentation of nucleotide and amino acid sequences in European patent applications*

If the subject-matter of the invention disclosed in a European patent application is nucleotide or amino acid sequences, the description shall contain a sequence listing complying with WIPO Standard ST. 25 reproduced in the **annex** to this decision.

**Article 2**

*Filing of sequence listings on electronic data carriers*

(1) In addition, a data carrier shall be filed containing the sequence listing and complying with the standard reproduced in the **annex** to this decision.

(2) If a sequence listing is corrected after being filed with the European Patent Office, a new data carrier containing the corrected sequence listing shall be filed.

(3) Each data carrier shall be accompanied by a statement by the applicant that the information recorded on the data carrier is identical to the written sequence listing.

**Article 3**

*Rectification of deficiencies*

(1) If the requirements laid down in Rule 27a EPC and Articles 1 and 2 of this decision are not met, the Receiving Section shall inform the applicant accordingly and invite him to remedy the disclosed deficiencies within a period of two months (Rule 41(1) EPC).

**Décision du Président de l'Office européen des brevets, en date du 2 octobre 1998, relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences**

Le Président de l'Office européen des brevets, vu l'article 10 CBE, les règles 27bis et 49 CBE ainsi que la règle 13<sup>ter</sup> PCT, décide :

**Article premier**

*Présentation normalisée du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet européen*

Si l'invention exposée dans une demande de brevet européen a pour objet des séquences de nucléotides ou d'acides aminés, la description doit contenir un listage des séquences établi conformément à la norme ST. 25 de l'OMPI, reproduite dans l'**annexe** à la présente décision.

**Article 2**

*Dépôt de listages des séquences sur support électronique de données*

(1) En outre, il y a lieu de déposer un support de données contenant un listage des séquences établi conformément à la norme reproduite dans l'**annexe** à la présente décision.

(2) Si un listage des séquences est rectifié après avoir été déposé auprès de l'Office européen des brevets, il y a lieu de déposer un nouveau support de données contenant le listage des séquences rectifié.

(3) Tout support de données doit être accompagné d'une déclaration du demandeur selon laquelle l'information figurant sur ce support est identique à celle que contient le listage écrit.

**Article 3**

*Correction d'irrégularités*

(1) S'il n'est pas satisfait aux exigences visées à la règle 27bis CBE et aux articles premier et 2 de la présente décision, la section de dépôt le signale au demandeur et l'invite à remédier aux irrégularités constatées dans un délai de deux mois (règle 41(1) CBE).

(2) Die Frist wird einmal um höchstens zwei Monate verlängert, sofern der Anmelder dies vor Ablauf der Frist beantragt.

(3) Ist ein Datenträger beschädigt oder aus einem anderen Grund nicht lesbar, so teilt die Eingangsstelle dies dem Anmelder mit und fordert ihn auf, einen neuen, einwandfrei lesbaren Datenträger innerhalb einer von ihr zu bestimmenden Frist einzureichen.

#### Artikel 4

*Einreichung von Sequenzprotokollen beim EPA als Internationaler Recherchenbehörde*

In bezug auf internationale Anmeldungen nach dem PCT, für die nach Regel 5.2 PCT ein Sequenzprotokoll vorgeschrieben ist, verfährt das Europäische Patentamt als Internationale Recherchenbehörde nach Regel 13<sup>er</sup>.1 PCT und Abschnitten 208 und 513 mit Anhang C der Verwaltungsvorschriften zum PCT.

#### Artikel 5

*Einreichung von Sequenzprotokollen beim EPA als Bestimmungsamt oder als ausgewähltes Amt*

Bei Aufforderung zur Einreichung eines Sequenzprotokolls nach Regel 104b (3a) EPÜ ist Artikel 3 entsprechend anzuwenden.

#### Artikel 6

*Veröffentlichung von Sequenzprotokollen durch das EPA*

Ein am Anmeldetag der europäischen Patentanmeldung eingereichtes Sequenzprotokoll wird als Bestandteil der Beschreibung zusammen mit den Anmeldungsunterlagen veröffentlicht. Ein nach dem Anmeldetag eingereichtes oder berichtigtes Sequenzprotokoll wird, wenn es vor dem Zeitpunkt vorliegt, an dem die technischen Vorbereitungen für die Veröffentlichung als abgeschlossen gelten, als Anlage zu den Anmeldungsunterlagen, andernfalls, bei Patenterteilung, als Anlage zur europäischen Patentschrift veröffentlicht.

(2) This period shall be extendable once only for a maximum of two months, provided the applicant requests such an extension prior to the expiry of the period.

(3) If a data carrier is damaged, or unreadable for some other reason, the Receiving Section shall inform the applicant accordingly and invite him to submit a flawless replacement within a period which it shall lay down.

#### Article 4

*Filing of sequence listings with the EPO as International Searching Authority*

With regard to international applications under the PCT, for which a sequence listing is prescribed under Rule 5.2 PCT, the European Patent Office shall act as International Searching Authority in accordance with Rule 13<sup>er</sup>.1 PCT and Sections 208 and 513 together with Annex C of the PCT Administrative Instructions.

#### Article 5

*Filing of sequence listings with the EPO as designated Office or elected Office*

Where an applicant is invited to file a sequence listing in accordance with Rule 104b(3a) EPC, Article 3 shall apply *mutatis mutandis*.

#### Article 6

*Publication of sequence listings by the EPO*

A sequence listing filed on the date of filing of the European patent application shall be published together with the application documents as part of the description. A sequence listing filed or corrected after the date of filing shall be published as an annex to the application documents, provided it is received before the date on which the technical preparations for publication are deemed to have been completed; otherwise, it shall be published as an annex to the European patent specification when the patent is granted.

(2) Le délai est prolongé une seule fois de deux mois au maximum, si le demandeur présente une requête en ce sens avant l'expiration du délai.

(3) Si un support de données est endommagé ou rendu illisible pour quelque autre raison, la section de dépôt le signale au demandeur et l'invite à déposer un nouveau support de données parfaitement lisible dans le délai qu'elle lui impartit.

#### Article 4

*Dépôt de listages des séquences auprès de l'OEB agissant en qualité d'administration chargée de la recherche internationale*

En ce qui concerne les demandes internationales déposées au titre du PCT pour lesquelles la règle 5.2 PCT prescrit un listage des séquences, l'Office européen des brevets, agissant en qualité d'administration chargée de la recherche internationale, procède conformément aux dispositions de la règle 13<sup>er</sup>.1 PCT et aux instructions administratives 208 et 513 du PCT, accompagnées de l'annexe C des Instructions administratives du PCT.

#### Article 5

*Dépôt de listages des séquences auprès de l'OEB agissant en tant qu'office désigné ou élu*

Lors de l'envoi d'une invitation à déposer un listage des séquences conformément à la règle 104ter (3bis) CBE, l'article 3 est applicable par analogie.

#### Article 6

*Publication par l'OEB de listages des séquences*

Un listage des séquences déposé à la date de dépôt de la demande de brevet européen est publié comme élément de la description avec les documents de la demande. Si un listage des séquences déposé ou rectifié après la date de dépôt est disponible avant la date à laquelle les préparatifs techniques entrepris en vue de la publication de la demande sont réputés achevés, il est publié en annexe aux documents de la demande, ou, si ce n'est pas le cas, lors de la délivrance du brevet, en annexe au fascicule du brevet européen.

**Artikel 7**

*Inkrafttreten*

Dieser Beschluß tritt am **1. Januar 1999** in Kraft und ersetzt den Beschluß vom 11. Dezember 1992, Beilage Nr. 2 zu ABI. EPA 12/1992, 1.

Geschehen zu München am  
2. Oktober 1998

*INGO KOBER*  
Präsident

**Article 7**

*Entry into force*

This decision shall enter into force on **1 January 1999** and shall replace the decision dated 11 December 1992, Supplement No. 2 to OJ EPO 12/1992, 1.

Done at Munich, 2 October 1998

*INGO KOBER*  
President

**Article 7**

*Entrée en vigueur*

La présente décision entre en vigueur le **1<sup>er</sup> janvier 1999** et remplace la décision du 11 décembre 1992, publiée dans le supplément n° 2 au Journal officiel 12/1992, 1.

Fait à Munich, le 2 octobre 1998

*INGO KOBER*  
Président

**ANHANG zum Beschluß des Präsidenten des Europäischen Patentamts vom 2. Oktober 1998 über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen**

**WIPO-Standard ST. 25**

**Standard für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in Patentanmeldungen**

Vom (ausführenden) PCIPI-Koordinierungsausschuß in seiner zweiundzwanzigsten Sitzung am 28. Mai 1998 angenommener Standard

Den Ämtern wird empfohlen, die Vorschriften des "Standards für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen nach dem Vertrag über die internationale Zusammenarbeit auf dem Gebiet des Patentwesens (PCT)", die in Anhang C zu den Verwaltungsvorschriften zum PCT enthalten sind, auch auf solche Patentanmeldungen anzuwenden, die keine internationalen Anmeldungen nach dem PCT sind, wobei zu beachten ist, daß bestimmte Vorschriften, die auf PCT-spezifische Verfahren und Erfordernisse zugeschnitten sind, auf Patentanmeldungen, die keine internationalen Anmeldungen nach dem PCT sind, möglicherweise keine Anwendung finden.\* Der Wortlaut des PCT-Standards ist nachfolgend wiedergegeben.

**ANNEX to the Decision of the President of the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings**

**WIPO Standard ST. 25**

**Standard for the presentation of nucleotide and amino acid sequence listings in patent applications**

Standard adopted by the PCIPI Executive Coordination Committee at its twenty-second session on May 28, 1998

It is recommended that Offices apply the provisions of the "Standard for the Presentation of Nucleotide and Amino Acid Sequence Listings in International Patent Applications Under the Patent Cooperation Treaty (PCT)" as set out in Annex C to the Administrative Instructions under the PCT, mutatis mutandis, to all patent applications other than the PCT international applications, noting that certain provisions specific to PCT procedures and requirements may not be applicable to patent applications other than PCT international applications.\* The text of that PCT Standard is reproduced on the following pages.

**ANNEXE à la décision du Président de l'Office européen des brevets, en date du 2 octobre 1998, relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences**

**Norme ST. 25 de l'OMPI**

**Norme relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet**

Norme adoptée par le Comité exécutif de coordination du PCIPI à sa vingt-deuxième session

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.\* Le texte de cette norme du PCT est reproduit ci-après.

\* Ist am 1. Juli 1998 die für ein Amt geltende nationale Gesetzgebung und Rechtspraxis nicht mit den Vorschriften der ersten beiden Sätze in Absatz 3 des "Standards für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen nach dem Vertrag über die internationale Zusammenarbeit auf dem Gebiet des Patentwesens (PCT)" vereinbar, so steht es diesem Amt frei, die betreffenden Vorschriften nicht zu befolgen, solange die Unvereinbarkeit besteht.

\* If, on July 1, 1998, the national law and practice applicable by an Office is not compatible with the provisions of the first two sentences of paragraph 3 of the "Standard for the Presentation of Nucleotide and Amino Acid Sequence Listings in International Patent Applications Under the Patent Cooperation Treaty (PCT)", that Office may choose not to follow those provisions for as long as that incompatibility continues.

\* Si, le 1<sup>er</sup> juillet 1998, la législation et la pratique nationales appliquées par un office ne sont pas compatibles avec les dispositions des deux premières phrases du paragraphe 3 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)", l'office peut choisir de ne pas appliquer ces dispositions tant que l'incompatibilité demeure.



## Anhang C zu den PCT-Verwaltungsvorschriften

Standard für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen nach dem PCT

### EINFÜHRUNG

1. Mit dem vorliegenden Standard soll auf eine standardisierte Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Patentanmeldungen hingewirkt werden. Anhand dieses Standards kann der Anmelder ein einheitliches Sequenzprotokoll erstellen, das von allen Anmeldeämtern, Internationalen Recherchenbehörden und mit der internationalen vorläufigen Prüfung beauftragten Behörden für die Zwecke der internationalen Phase und von allen Bestimmungsämtern und ausgewählten Ämtern für die Zwecke der nationalen Phase akzeptiert wird. Der Standard soll ferner eine genauere und qualitativ bessere Darstellung von Nucleotid- und Aminosäuresequenzen in internationalen Anmeldungen sicherstellen, die Darstellung und Verbreitung der Sequenzen zum Nutzen der Anmelder, der Öffentlichkeit und der Prüfer erleichtern, die Recherche von Sequenzdaten vereinfachen und den Austausch dieser Daten in elektronischer Form sowie ihre Aufnahme in computergestützte Datenbanken ermöglichen.

### DEFINITIONEN

2. Im Rahmen dieses Standards gelten folgende Definitionen:

i) Unter einem "Sequenzprotokoll" ist ein Teil der Beschreibung der Anmeldung in der eingereichten Fassung oder ein zur Anmeldung nachgereichtes Schriftstück zu verstehen, das die Nucleotid- und/oder Aminosäuresequenzen im einzelnen offenbart und sonstige verfügbare Angaben enthält.

ii) In das Protokoll dürfen nur unverzweigte Sequenzen von mindestens 4 Aminosäuren bzw. unverzweigte Sequenzen von mindestens 10 Nucleotiden aufgenommen werden. Verzweigte Sequenzen, Sequenzen mit weniger als 4 genau definierten Nucleotiden oder Aminosäuren sowie Sequenzen mit Nucleotiden oder Aminosäuren, die nicht in Anhang 2, Tabellen 1, 2, 3 und 4 aufgeführt sind, sind ausdrücklich ausgenommen.

iii) Unter "Nucleotiden" sind nur Nucleotide zu verstehen, die mittels der in Anhang 2, Tabelle 1 aufge-

## Annex C to the PCT Administrative Instructions

Standard for the presentation of nucleotide and amino acid sequence listings in international patent applications under the PCT

### INTRODUCTION

1. This Standard has been elaborated so as to provide standardization of the presentation of nucleotide and amino acid sequence listings in international patent applications. The Standard is intended to allow the applicant to draw up a single sequence listing which is acceptable to all receiving Offices, International Searching and Preliminary Examining Authorities for the purposes of the international phase, and to all designated and elected Offices for the purposes of the national phase. It is intended to enhance the accuracy and quality of presentations of nucleotide and amino acid sequences given in international applications, to make for easier presentation and dissemination of sequences for the benefit of applicants, the public and examiners, to facilitate searching of sequence data and to allow the exchange of sequence data in electronic form and the introduction of sequence data onto computerized databases.

### DEFINITIONS

2. For the purposes of this Standard:

(i) the expression "sequence listing" means a part of the description of the application as filed or a document filed subsequently to the application, which gives a detailed disclosure of the nucleotide and/or amino acid sequences and other available information;

(ii) sequences which are included are any unbranched sequences of four or more amino acids or unbranched sequences of ten or more nucleotides. Branched sequences, sequences with fewer than four specifically defined nucleotides or amino acids as well as sequences comprising nucleotides or amino acids other than those listed in Appendix 2, Tables 1, 2, 3 and 4, are specifically excluded from this definition;

(iii) "nucleotides" embrace only those nucleotides that can be represented using the symbols set forth

## Annexe C des Instructions administratives du PCT

Norme relative à la présentation du listage 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 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 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

fürten Symbole wiedergegeben werden können. Modifikationen wie z. B. methylierte Basen können nach der Anleitung in Anhang 2, Tabelle 2 beschrieben werden, sind aber in der Nucleotidsequenz nicht explizit auszuweisen.

iv) Unter "Aminosäuren" sind die in Anhang 2, Tabelle 3 aufgeführten gängigen L-Aminosäuren aus den natürlich vorkommenden Proteinen zu verstehen. Aminosäuresequenzen, die mindestens eine D-Aminosäure enthalten, fallen nicht unter diese Definition. Aminosäuresequenzen, die posttranslational modifizierte Aminosäuren enthalten, können mittels der in Anhang 2, Tabelle 3 aufgeführten Symbole wie die ursprünglich translatierte Aminosäure beschrieben werden, wobei die modifizierten Positionen wie z. B. Hydroxylierungen oder Glykosilierungen nach der Anleitung in Anhang 2, Tabelle 4 beschrieben werden, diese Modifikationen aber in der Aminosäuresequenz nicht explizit auszuweisen sind. Unter die vorstehende Definition fallen auch Peptide oder Proteine, die anhand der in Anhang 2, Tabelle 3 aufgeführten Symbole sowie einer an anderer Stelle aufgenommenen Beschreibung, die beispielsweise Aufschluß über ungewöhnliche Bindungen, Quervernetzungen (z. B. Disulfidbrücken) und "end caps", Nichtpeptidbindungen usw. gibt, als Sequenz wiedergegeben werden können.

v) Unter "Sequenzkennzahl" ist eine Zahl zu verstehen, die jeder im Protokoll aufgeführten Sequenz als SEQ ID NO zugewiesen wird.

vi) Die "numerische Kennzahl" ist eine dreistellige Zahl, die für ein bestimmtes Datenelement steht.

vii) Unter "sprachneutralem Vokabular" ist ein festes Vokabular zu verstehen, das im Sequenzprotokoll zur Wiedergabe der vom Hersteller einer Sequenzdatenbank vorgeschriebenen wissenschaftlichen Begriffe verwendet wird (dazu gehören wissenschaftliche Namen, nähere Bestimmungen und ihre Entsprechungen im festen Vokabular, die Symbole in Anhang 2, Tabellen 1, 2, 3 und 4 und die Merkmalschlüssel in Anhang 2, Tabellen 5 und 6).

viii) Unter "zuständiger Behörde" ist die Internationale Recherchenbehörde oder die mit der internationalen vorläufigen Prüfung beauftragte Behörde, die die internationale Recherche bzw. die internationale

in Appendix 2, Table 1. Modifications, for example, methylated bases, may be described as set forth in Appendix 2, Table 2, but shall not be shown explicitly in the nucleotide sequence;

(iv) "amino acids" are those L-amino acids commonly found in naturally occurring proteins and are listed in Appendix 2, Table 3. Those amino acid sequences containing at least one D-amino acid are not intended to be embraced by this definition. Any amino acid sequence that contains post-translationally modified amino acids may be described as the amino acid sequence that is initially translated using the symbols shown in Appendix 2, Table 3, with the modified positions, for example, hydroxylations or glycosylations, being described as set forth in Appendix 2, Table 4, but these modifications shall not be shown explicitly in the amino acid sequence. Any peptide or protein that can be expressed as a sequence using the symbols in Appendix 2, Table 3, in conjunction with a description elsewhere to describe, for example, abnormal linkages, cross-links (for example, disulfide bridge) and end caps, non-peptidyl bonds, etc., is embraced by this definition;

(v) "sequence identifier" is a unique integer that corresponds to the SEQ ID NO assigned to each sequence in the listing;

(vi) "numeric identifier" is a three-digit number which represents a specific data element;

(vii) "language-neutral vocabulary" is a controlled vocabulary used in the sequence listing that represents scientific terms as prescribed by sequence database providers (including scientific names, qualifiers and their controlled-vocabulary values, the symbols appearing in Appendix 2, Tables 1, 2, 3 and 4, and the feature keys appearing in Appendix 2, Tables 5 and 6);

(viii) "competent Authority" is the International Searching Authority that is to carry out the international search on the international application, or the International Preliminary Examining Authority that is to carry

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 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 de la recherche internationale ou l'administration chargée de l'examen préliminaire international pour la demande internationale

vorläufige Prüfung zu der internationalen Anmeldung durchführt, oder das Bestimmungsamt bzw. ausgewählte Amt zu verstehen, das mit der Bearbeitung der internationalen Anmeldung begonnen hat.

#### SEQUENZPROTOKOLL

3. Das Sequenzprotokoll im Sinne der Nummer 2 i) ist ans Ende der Anmeldung zu setzen, wenn es mit ihr zusammen eingereicht wird. Dieser Teil ist mit "Sequenzprotokoll" zu überschreiben, muß mit einer neuen Seite beginnen und sollte gesondert numeriert werden. Das Sequenzprotokoll ist Bestandteil der Beschreibung; vorbehaltlich Nummer 36 erübrigt es sich deshalb, die Sequenzen in der Beschreibung an anderer Stelle nochmals zu beschreiben.

4. Wird das Sequenzprotokoll im Sinne der Nummer 2 i) nicht zusammen mit der Anmeldung eingereicht, sondern als gesondertes Schriftstück nachgereicht (s. Nr. 37), so ist es mit der Überschrift "Sequenzprotokoll" und einer gesonderten Seitennummerierung zu versehen. Die in der Anmeldung in der eingereichten Fassung gewählte Numerierung der Sequenzen (s. Nr. 5) ist auch im nachgereichten Sequenzprotokoll beizubehalten.

5. Jeder Sequenz wird eine eigene Sequenzkennzahl zugeteilt. Die Kennzahlen beginnen mit 1 und setzen sich in aufsteigender Reihenfolge als ganze Zahlen fort. Gibt es zu einer Kennzahl keine Sequenz, so ist am Anfang der auf die SEQ ID NO folgenden Zeile unter der numerischen Kennzahl <400> der Code 000 anzugeben. Unter der numerischen Kennzahl <160> ist die Gesamtzahl der SEQ ID NOs anzugeben, und zwar unabhängig davon, ob im Anschluß an eine SEQ ID NO eine Sequenzkennzahl oder der Code 000 folgt.

6. In der Beschreibung, in den Ansprüchen und in den Zeichnungen der Anmeldung ist auf die im Sequenzprotokoll dargestellten Sequenzen mit "SEQ ID NO", gefolgt von der betreffenden Kennzahl, zu verweisen.

7. Für die Darstellung von Nucleotid- und Aminosäuresequenzen ist zumindest eine der folgenden drei Möglichkeiten zu wählen:

- i) nur Nucleotidsequenz
- ii) nur Aminosäuresequenz
- iii) Nucleotidsequenz zusammen mit der entsprechenden Aminosäuresequenz

out the international preliminary examination on the international application, or the designated/elected Office before which the processing of the international application has started.

#### SEQUENCE LISTING

3. The sequence listing as defined in paragraph 2(i) shall, where it is filed together with the application, be placed at the end of the application. This part shall be entitled "Sequence Listing," begin on a new page and preferably have independent page numbering. The sequence listing forms an integral part of the description; it is therefore unnecessary, subject to paragraph 36, to describe the sequences elsewhere in the description.

4. Where the sequence listing as defined in paragraph 2(i) is not contained in the application as filed but is a separate document furnished subsequently to the filing of the application (see paragraph 37), it shall be entitled "Sequence Listing" and shall have independent page numbering. The original numbering of the sequences (see paragraph 5) in the application as filed shall be maintained in the subsequently furnished sequence listing.

5. Each sequence shall be assigned a separate sequence identifier. The sequence identifiers shall begin with 1 and increase sequentially by integers. If no sequence is present for a sequence identifier, the code 000 should appear under numeric identifier <400>, beginning on the next line following the SEQ ID NO. The response for numeric identifier <160> shall include the total number of SEQ ID NOs, whether followed by a sequence or by the code 000.

6. In the description, claims or drawings of the application, the sequences represented in the sequence listing shall be referred to by the sequence identifier and preceded by "SEQ ID NO:".

7. Nucleotide and amino acid sequences should be represented by at least one of the following three possibilities:

- (i) a pure nucleotide sequence;
- (ii) a pure amino acid sequence;
- (iii) a nucleotide sequence together with its corresponding amino acid sequence.

en question, ou encore l'office désigné ou élu au sein duquel le traitement de la demande internationale a commencé.

#### LISTAGE DES SÉQUENCES

3. Le listage des séquences défini au paragraphe 2.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 "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 36, de décrire les séquences ailleurs dans la description.

4. Lorsque le listage des séquences défini au paragraphe 2.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 37), il doit s'intituler "Listage des séquences" et faire l'objet d'une pagination distincte. La numérotation des séquences (voir le paragraphe 5) figurant dans la demande telle qu'elle a été déposée doit être maintenue dans un listage des séquences remis ultérieurement.

5. A 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.

Bei einer Offenbarung im unter Ziffer iii genannten Format muß die Aminosäuresequenz als solche im Sequenzprotokoll mit einer eigenen Sequenzkennzahl gesondert offenbart werden.

For those sequences disclosed in the format specified in option (iii), above, the amino acid sequence must be disclosed separately in the sequence listing as a pure amino acid sequence with a separate integer sequence identifier.

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.

## NUCLEOTIDSEQUENZEN

### *Zu verwendende Symbole*

8. Nucleotidsequenzen sind nur anhand eines Einzelstrangs in Richtung vom 5'-Ende zum 3'-Ende von links nach rechts wiederzugeben. Die Begriffe 3' und 5' werden in der Sequenz nicht dargestellt.

9. Die Basen einer Nucleotidsequenz sind anhand der einbuchstabigen Codes für Nucleotidsequenzzeichen darzustellen. Es dürfen nur die in Anhang 2, Tabelle 1 aufgeführten Kleinbuchstaben verwendet werden.

10. Modifizierte Basen sind in der Sequenz selbst wie die entsprechenden unmodifizierten Basen oder mit "n" wiederzugeben, wenn die modifizierte Base zu den in Anhang 2, Tabelle 2 aufgeführten gehört; die Modifikation ist im Merkmalsteil des Sequenzprotokolls anhand der Codes in Anhang 2, Tabelle 2 näher zu beschreiben. Diese Codes dürfen in der Beschreibung oder im Merkmalsteil des Sequenzprotokolls, nicht jedoch in der Sequenz selbst verwendet werden (s. auch Nr. 32). Das Symbol "n" steht immer nur für ein einziges unbekanntes oder modifiziertes Nucleotid.

### *Zu verwendendes Format*

11. Bei Nucleotidsequenzen sind höchstens 60 Basen pro Zeile – mit einem Leerraum zwischen jeder Gruppe von 10 Basen – aufzuführen.

12. Die Basen einer Nucleotidsequenz (einschließlich Intronen) sind jeweils in Zehnergruppen aufzuführen; dies gilt nicht für die codierenden Teile der Sequenz. Bleiben am Ende nicht-codierender Teile einer Sequenz weniger als 10 Basen übrig, so sind sie zu einer Gruppe zusammenzufassen und durch einen Leerraum von angrenzenden Gruppen zu trennen.

13. Die Basen der codierenden Teile einer Nucleotidsequenz sind als Triplets (Codonen) aufzuführen.

## NUCLEOTIDE SEQUENCES

### *Symbols to Be Used*

8. A nucleotide sequence shall be presented only by a single strand, in the 5'-end to 3'-end direction from left to right. The terms 3' and 5' shall not be represented in the sequence.

9. The bases of a nucleotide sequence shall be represented using the one-letter code for nucleotide sequence characters. Only lower case letters in conformity with the list given in Appendix 2, Table 1, shall be used.

10. Modified bases shall be represented as the corresponding unmodified bases or as "n" in the sequence itself if the modified base is one of those listed in Appendix 2, Table 2, and the modification shall be further described in the feature section of the sequence listing, using the codes given in Appendix 2, Table 2. These codes may be used in the description or the feature section of the sequence listing but not in the sequence itself (see also paragraph 32). The symbol "n" is the equivalent of only one unknown or modified nucleotide.

### *Format to Be Used*

11. A nucleotide sequence shall be listed with a maximum of 60 bases per line, with a space between each group of 10 bases.

12. The bases of a nucleotide sequence (including introns) shall be listed in groups of 10 bases, except in the coding parts of the sequence. Leftover bases, fewer than 10 in number at the end of non-coding parts of a sequence, should be grouped together and separated from adjacent groups by a space.

13. The bases of the coding parts of a nucleotide sequence shall be listed as triplets (codons).

## 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. Die Zählung der Nucleotidbasen beginnt bei der ersten Base der Sequenz mit 1. Von hier aus ist die gesamte Sequenz in 5'-3'-Richtung fortlaufend durchzuzählen. Am rechten Rand ist jeweils neben der Zeile mit den einbuchstabigen Codes für die Basen die Nummer der letzten Base dieser Zeile anzugeben. Die vorstehend beschriebene Zählweise für Nucleotidsequenzen gilt auch für Nucleotidsequenzen mit ringförmiger Konfiguration, wobei allerdings die Bestimmung des ersten Nucleotids der Sequenz dem Anmelder überlassen bleibt.

15. Eine Nucleotidsequenz, die aus einem oder mehreren nichtbenachbarten Abschnitten einer größeren Sequenz oder aus Abschnitten verschiedener Sequenzen besteht, ist als gesonderte Sequenz mit eigener Sequenzkennzahl zu numerieren. Sequenzen mit einer oder mehreren Lücken sind als mehrere gesonderte Sequenzen mit eigenen Sequenzkennzahlen zu numerieren, wobei die Zahl der gesonderten Sequenzen der Zahl der jeweils zusammenhängenden Sequenzdatenreihen entspricht.

#### AMINOSÄURESEQUENZEN

##### *Zu verwendende Symbole*

16. Die Aminosäuren einer Protein- oder Peptidsequenz sind in Richtung von der Amino- zur Carboxylgruppe von links nach rechts aufzuführen, wobei die Amino- und Carboxylgruppen in der Sequenz nicht darzustellen sind.

17. Die Aminosäuren sind anhand des dreibuchstabigen Codes mit großem Anfangsbuchstaben entsprechend der Liste in Anhang 2, Tabelle 3 darzustellen. Eine Aminosäuresequenz, die einen Leerraum oder interne Terminatorsymbole (z. B. "Ter", "\*" oder ".") enthält, ist nicht als eine einzige Aminosäuresequenz, sondern als getrennte Aminosäuresequenzen darzustellen (s. Nr. 22).

18. Modifizierte und seltene Aminosäuren sind in der Sequenz selbst wie die entsprechenden unmodifizierten Aminosäuren oder mit "Xaa" wiederzugeben, wenn sie zu den in Anhang 2, Tabelle 4 aufgeführten gehören und die Modifikation im Merkmalsteil des Sequenzprotokolls anhand der Codes in Anhang 2, Tabelle 4 näher beschrieben wird. Diese Codes dürfen in der Beschreibung oder im Merkmalsteil des

14. The enumeration of the nucleotide shall start at the first base of the sequence with number 1. It shall be continuous through the whole sequence in the direction 5' to 3'. It shall be marked in the right margin, next to the line containing the one-letter codes for the bases, and giving the number of the last base of that line. The enumeration method for nucleotide sequences set forth above remains applicable to nucleotide sequences that are circular in configuration, with the exception that the designation of the first nucleotide of the sequence may be made at the option of the applicant.

15. A nucleotide sequence that is made up of one or more non-contiguous segments of a larger sequence or of segments from different sequences shall be numbered as a separate sequence, with a separate sequence identifier. A sequence with a gap or gaps shall be numbered as a plurality of separate sequences with separate sequence identifiers, with the number of separate sequences being equal in number to the number of continuous strings of sequence data.

#### AMINO ACID SEQUENCES

##### *Symbols to Be Used*

16. The amino acids in a protein or peptide sequence shall be listed in the amino to carboxy direction from left to right. The amino and carboxy groups shall not be represented in the sequence.

17. The amino acids shall be represented using the three-letter code with the first letter as a capital and shall conform to the list given in Appendix 2, Table 3. An amino acid sequence that contains a blank or internal terminator symbols (for example, "Ter" or "\*" or ".") may not be represented as a single amino acid sequence, but shall be presented as separate amino acid sequences (see paragraph 22).

18. Modified and unusual amino acids shall be represented as the corresponding unmodified amino acids or as "Xaa" in the sequence itself if the modified amino acid is one of those listed in Appendix 2, Table 4, and the modification shall be further described in the feature section of the sequence listing, using the codes given in Appendix 2, Table 4. These codes may be used in the description or the feature section of

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

Sequenzprotokolls, nicht jedoch in der Sequenz selbst verwendet werden (s. auch Nr. 32). Das Symbol "Xaa" steht immer nur für eine einzige unbekannte oder modifizierte Aminosäure.

the sequence listing but not in the sequence itself (see also paragraph 32). The symbol "Xaa" is the equivalent of only one unknown or modified amino acid.

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.

#### *Zu verwendendes Format*

19. Bei Protein- oder Peptidsequenzen sind höchstens 16 Aminosäuren pro Zeile mit einem Leerraum zwischen den einzelnen Aminosäuren aufzuführen.

#### *Format to Be Used*

19. A protein or peptide sequence shall be listed with a maximum of 16 amino acids per line, with a space provided between each amino acid.

#### *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. Die den Codonen der codierenden Teile einer Nucleotidsequenz entsprechenden Aminosäuren sind unmittelbar unter den jeweiligen Codonen anzugeben. Wird ein Codon durch ein Intron aufgespalten, so ist das Aminosäuresymbol unter dem Teil des Codons anzugeben, der zwei Nucleotide enthält.

20. Amino acids corresponding to the codons in the coding parts of a nucleotide sequence shall be placed immediately under the corresponding codons. Where a codon is split by an intron, the amino acid symbol should be given below the portion of the codon containing two nucleotides.

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. Die Zählung der Aminosäuren beginnt bei der ersten Aminosäure der Sequenz mit 1. Fakultativ können die dem reifen Protein vorausgehenden Aminosäuren wie beispielsweise Präsequenzen, Prosequenzen, Präprosequenzen und Signalsequenzen, soweit vorhanden, mit negativem Vorzeichen numeriert werden, wobei die Rückwärtszählung mit der Aminosäure vor Nummer 1 beginnt. Null (0) wird nicht verwendet, wenn Aminosäuren zur Abgrenzung gegen das reife Protein mit negativem Vorzeichen numeriert werden. Die Nummern sind im Abstand von jeweils 5 Aminosäuren unter der Sequenz anzugeben. Die vorstehend beschriebene Zählweise für Aminosäuresequenzen gilt auch für Aminosäuresequenzen mit ringförmiger Konfiguration, wobei allerdings die Bestimmung der ersten Aminosäure der Sequenz dem Anmelder überlassen bleibt.

21. The enumeration of amino acids shall start at the first amino acid of the sequence, with number 1. Optionally, the amino acids preceding the mature protein, for example pre-sequences, pro-sequences, pre-pro-sequences and signal sequences, when present, may have negative numbers, counting backwards starting with the amino acid next to number 1. Zero (0) is not used when the numbering of amino acids uses negative numbers to distinguish the mature protein. It shall be marked under the sequence every five amino acids. The enumeration method for amino acid sequences set forth above remains applicable for amino acid sequences that are circular in configuration, with the exception that the designation of the first amino acid of the sequence may be made at the option of the applicant.

21. L'énumération des acides aminés doit commencer par le premier acide aminé de la séquence, qui portera le numéro 1. A titre facultatif, les acides aminés précédant la protéine mature, par exemple les pré-sequences, les prosequences et les pré-prosequences 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. Eine Aminosäuresequenz, die aus einem oder mehreren nichtbenachbarten Abschnitten einer größeren Sequenz oder aus Abschnitten verschiedener Sequenzen besteht, ist als gesonderte Sequenz mit eigener Sequenzkennzahl zu numerieren. Sequenzen mit einer oder mehreren Lücken sind als mehrere gesonderte Sequenzen mit eigenen Sequenzkennzahlen zu numerieren, wobei die Zahl der gesonderten Sequenzen der Zahl der jeweils zusammenhängenden Sequenzdatenreihen entspricht.

22. An amino acid sequence that is made up of one or more non-contiguous segments of a larger sequence or of segments from different sequences shall be numbered as a separate sequence, with a separate sequence identifier. A sequence with a gap or gaps shall be numbered as a plurality of separate sequences with separate sequence identifiers, with the number of separate sequences being equal in number to the number of continuous strings of sequence data.

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.

SONSTIGE VERFÜGBARE ANGABEN  
IM SEQUENZPROTOKOLL

23. Die Angaben sind im Sequenzprotokoll in der Reihenfolge anzugeben, in der sie in der Liste der numerischen Kennzahlen der Datenelemente in Anhang 1 aufgeführt sind.

24. Für die Angaben im Sequenzprotokoll sind nur die in Anhang 1 aufgeführten numerischen Kennzahlen, nicht aber die dazugehörigen Beschreibungen zu verwenden. Die Angaben müssen unmittelbar auf die numerische Kennzahl folgen; im Sequenzprotokoll brauchen nur diejenigen numerischen Kennzahlen angegeben zu werden, zu denen auch Angaben vorliegen. Die einzigen beiden Ausnahmen zu dieser Vorschrift bilden die numerischen Kennzahlen <220> und <300>, die für die Rubrik "Merkmal" bzw. "Veröffentlichungsangaben" stehen und mit den Angaben unter den numerischen Kennzahlen <221> bis <223> bzw. <301> bis <313> zusammenhängen. Werden unter diesen numerischen Kennzahlen im Sequenzprotokoll Angaben zu den Merkmalen oder zur Veröffentlichung gemacht, so sollte auch die numerische Kennzahl <220> bzw. <300> aufgeführt, die dazugehörige Rubrik aber nicht ausgefüllt werden. Generell sollte zwischen den numerischen Kennzahlen eine Leerzeile eingefügt werden, wenn sich die an erster oder zweiter Position der numerischen Kennzahl stehende Ziffer ändert. Eine Ausnahme von dieser allgemeinen Regel bildet die numerische Kennzahl <310>, der keine Leerzeile vorausgehen darf. Auch vor jeder Wiederholung der numerischen Kennzahl ist eine Leerzeile einzufügen.

*Obligatorische Datenelemente*

25. In das Sequenzprotokoll sind außerdem vor der eigentlichen Nucleotid- und/oder Aminosäuresequenz die folgenden in Anhang 1 definierten Angaben (obligatorische Datenelemente) aufzunehmen:

<110> Name des Anmelders  
<120> Bezeichnung der Erfindung  
<160> Anzahl der SEQ ID NOs  
<210> SEQ ID NO: x  
<211> Länge  
<212> Art  
<213> Organismus  
<400> Sequenz

OTHER AVAILABLE INFORMATION IN  
THE SEQUENCE LISTING

23. The order of the items of information in the sequence listings shall follow the order in which those items are listed in the list of numeric identifiers of data elements as defined in Appendix 1.

24. Only numeric identifiers of data elements as defined in Appendix 1 shall be used for the presentation of the items of information in the sequence listing. The corresponding numeric identifier descriptions shall not be used. The provided information shall follow immediately after the numeric identifier while only those numeric identifiers for which information is given need appear on the sequence listing. Two exceptions to this requirement are numeric identifiers <220> and <300>, which serve as headers for "Feature" and "Publication Information," respectively, and are associated with information in numeric identifiers <221> to <223> and <301> to <313>, respectively. When feature and publication information is provided in the sequence listing under those numeric identifiers, numeric identifiers <220> and <300>, respectively, should be included, but left blank. Generally, a blank line shall be inserted between numeric identifiers when the digit in the first or second position of the numeric identifier changes. An exception to this general rule is that no blank line should appear preceding numeric identifier <310>. Additionally, a blank line shall precede any repeated numeric identifier.

*Mandatory Data Elements*

25. The sequence listing shall include, in addition to and immediately preceding the actual nucleotide and/or amino acid sequence, the following items of information defined in Appendix 1 (mandatory data elements):

<110> Applicant name  
<120> Title of invention  
<160> Number of SEQ ID NOs  
<210> SEQ ID NO: x  
<211> Length  
<212> Type  
<213> Organism  
<400> Sequence

AUTRES RENSEIGNEMENTS  
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.

*Eléments 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

Ist der Name des Anmelders (numerische Kennzahl <110>) nicht in Buchstaben des lateinischen Alphabets geschrieben, so ist er – im Wege der Transliteration oder der Übersetzung ins Englische – auch in lateinischen Buchstaben anzugeben.

Where the name of the applicant (numeric identifier <110>) is written in characters other than those of the Latin alphabet, it shall also be indicated in characters of the Latin alphabet either as a mere transliteration or through translation into English.

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.

Die Datenelemente mit Ausnahme der Angaben unter den numerischen Kennzahlen <110>, <120> und <160> sind für jede im Sequenzprotokoll aufgeführte Sequenz zu wiederholen. Gibt es zu einer Sequenzkennzahl keine Sequenz, so müssen nur die Datenelemente unter den numerischen Kennzahlen <210> und <400> angegeben werden (s. Nr. 5 und SEQ ID NO: 4 in dem in Anhang 3 dieses Standards enthaltenen Beispiel).

The data elements, except those under numeric identifiers <110>, <120> and <160>, shall be repeated for each sequence included in the sequence listing. Only the data elements under numeric identifiers <210> and <400> are mandatory if no sequence is present for a sequence identifier (see paragraph 5, above, and SEQ ID NO: 4 in the example depicted in Appendix 3 of this Standard).

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. In Sequenzprotokolle, die zusammen mit der dazugehörigen Anmeldung oder vor Vergabe einer Anmeldeungsnummer eingereicht werden, ist neben den unter Nummer 25 genannten Datenelementen auch das folgende aufzunehmen:

26. In addition to the data elements identified in paragraph 25, above, when a sequence listing is filed at the same time as the application to which it pertains or at any time prior to the assignment of an application number, the following data element shall be included in the sequence listing:

26. Outre les éléments de données indiqués au paragraphe 25, 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> Aktenzeichen

<130> File reference

<130> Référence du dossier

27. In Sequenzprotokolle, die auf Anforderung einer zuständigen Behörde oder nach Vergabe einer Anmeldeungsnummer eingereicht werden, sind neben den unter Nummer 25 genannten Datenelementen auch die folgenden aufzunehmen:

27. In addition to the data elements identified in paragraph 25, above, when a sequence listing is filed in response to a request from a competent Authority or at any time following the assignment of an application number, the following data elements shall be included in the sequence listing:

27. Outre les éléments de données énumérés au paragraphe 25, 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> Vorliegende Patentanmeldung  
<141> Anmeldetag der vorliegenden Anmeldung

<140> Current patent application  
<141> Current filing date

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

28. In Sequenzprotokolle, die zu einer Anmeldung eingereicht werden, die die Priorität einer früheren Anmeldung in Anspruch nimmt, sind neben den unter Nummer 25 genannten Datenelementen auch die folgenden aufzunehmen:

28. In addition to the data elements identified in paragraph 25, above, when a sequence listing is filed relating to an application which claims the priority of an earlier application, the following data elements shall be included in the sequence listing:

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> Frühere Patentanmeldung

<150> Earlier patent application

<150> Demande de brevet antérieure

<151> Anmeldetag der früheren Anmeldung

<151> Earlier application filing date

<151> Date de dépôt de la demande antérieure



29. Wird in der Sequenz "n", "Xaa" oder eine modifizierte Base oder modifizierte/seltene L-Aminosäure aufgeführt, so müssen die folgenden Datenelemente angegeben werden:

<220> Merkmal  
<221> Name/Schlüssel  
<222> Lage  
<223> Sonstige Informationen

30. Ist der Organismus (numerische Kennzahl <213>) eine "künstliche Sequenz" oder "unbekannt", so müssen die folgenden Datenelemente angegeben werden:

<220> Merkmal  
<223> Sonstige Angaben

#### *Fakultative Datenelemente*

31. Alle in Anhang 1 definierten Datenelemente, die unter den Nummern 25 bis 30 nicht erwähnt sind, sind fakultativ (fakultative Datenelemente).

#### *Angabe von Merkmalen*

32. Merkmale, die (unter der numerischen Kennzahl <220>) zu einer Sequenz angegeben werden, sind durch die in Anhang 2, Tabellen 5 und 6<sup>1</sup> aufgeführten "Merkmal-schlüssel" zu beschreiben.

#### *Freier Text*

33. Unter "freiem Text" ist eine verbale Beschreibung der Eigenschaften der Sequenz ohne Verwendung des sprachneutralen Vokabulars im Sinne der Nummer 2 (vii) unter der numerischen Kennzahl <223> (Sonstige Angaben) zu verstehen.

34. Der freie Text sollte sich auf einige kurze, für das Verständnis der Sequenz unbedingt notwendige Begriffe beschränken. Er sollte in englischer Sprache für jedes Datenelement nicht länger als 4 Zeilen sein und höchstens 65 Buchstaben pro Zeile umfassen. Alle weiteren Angaben sind in den Hauptteil der Beschreibung in der dort verwendeten Sprache aufzunehmen.

35. Freier Text sollte in englischer Sprache abgefaßt sein.

36. Enthält das Sequenzprotokoll, das Bestandteil der Beschreibung ist, freien Text, so ist dieser im Hauptteil der Beschreibung in der dort verwen-

29. If "n" or "Xaa" or a modified base or modified/unusual L-amino acid is used in the sequence, the following data elements are mandatory:

<220> Feature  
<221> Name/key  
<222> Location  
<223> Other information

30. If the organism (numeric identifier <213>) is "Artificial Sequence" or "Unknown," the following data elements are mandatory:

<220> Feature  
<223> Other information

#### *Optional Data Elements*

31. All data elements defined in Appendix 1, not mentioned in paragraphs 25 to 30, above, are optional (optional data elements).

#### *Presentation of Features*

32. When features of sequences are presented (that is, numeric identifier <220>), they shall be described by the "feature keys" set out in Appendix 2, Tables 5 and 6.<sup>1</sup>

#### *Free Text*

33. "Free text" is a wording describing characteristics of the sequence under numeric identifier <223> (Other information) which does not use language-neutral vocabulary as referred to in paragraph 2(vii).

34. The use of free text shall be limited to a few short terms indispensable for the understanding of the sequence. It shall not exceed four lines with a maximum of 65 characters per line for each given data element, when written in English. Any further information shall be included in the main part of the description in the language thereof.

35. Any free text should preferably be in the English language.

36. Where the sequence listing part of the description contains free text, any such free text shall be repeated in the main part of the description in

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

#### *Eléments 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.

#### *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.

36. Lorsque, dans la description, la partie réservée au listage des séquences contient du texte libre, celui-ci doit être répété dans la partie

<sup>1</sup> Diese Tabellen enthalten Auszüge aus den Merkmals-tabellen "DDBJ/EMBL/GenBank Feature Table" (Nucleotidsequenzen) und "SWISS PROT Feature Table" (Aminosäuresequenzen).

<sup>1</sup> These tables contain extracts of the DDBJ/EMBL/GenBank Feature Table (nucleotide sequences) and the SWISS PROT Feature Table (amino acid sequences).

<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).

deten Sprache zu wiederholen. Es wird empfohlen, den in der Sprache des Hauptteils der Beschreibung abgefaßten freien Text in einen besonderen Abschnitt der Beschreibung mit der Überschrift "Sequenzprotokoll – freier Text" aufzunehmen.

#### NACHGERECHTES SEQUENZPROTOKOLL

37. Alle Sequenzprotokolle, die nicht zusammen mit der Anmeldung, sondern später eingereicht werden, dürfen nicht über den Offenbarungsgehalt der Anmeldung in der eingereichten Fassung hinausgehen; ihnen ist eine entsprechende Erklärung beizufügen. Die zur Anmeldung nachgereichten Sequenzprotokolle dürfen also nur die Sequenzen enthalten, die in der Anmeldung in der eingereichten Fassung offenbart worden sind.

38. Sequenzprotokolle, die nicht zusammen mit der Anmeldung eingereicht worden sind, sind nicht Bestandteil der Anmeldung. Da die Regeln 13<sup>ter</sup>, 26.3 und 91 sowie der Artikel 34 PCT Anwendung finden, besteht jedoch – vorbehaltlich der einschlägigen Bestimmungen – die Möglichkeit, Sequenzprotokolle, die zusammen mit der Anmeldung eingereicht wurden, im Wege der Mängelbeseitigung gemäß Regel 13<sup>ter</sup> oder 26.3 PCT und – bei Vorliegen offensichtlicher Fehler – gemäß Regel 91 PCT zu berichtigen oder gemäß Artikel 34 PCT zu ändern sowie ferner im Wege der Änderung der Anmeldung gemäß Artikel 34 PCT ein Sequenzprotokoll nachzureichen.

#### SEQUENZPROTOKOLL IN MASCHINENLESBARER FORM

39. Das in der Anmeldung enthaltene Sequenzprotokoll ist zusätzlich auch in maschinenlesbarer Form einzureichen, wenn die zuständige Behörde dies verlangt.

40. Das zusätzlich eingereichte maschinenlesbare Sequenzprotokoll muß mit dem geschriebenen Protokoll identisch sein und ist zusammen mit einer Erklärung folgenden Wortlauts einzureichen: "Die in maschinenlesbarer Form aufgezeichneten Angaben sind mit dem geschriebenen Sequenzprotokoll identisch."

the language thereof. It is recommended that the free text in the language of the main part of the description be put in a specific section of the description called "Sequence Listing Free Text."

#### SUBSEQUENTLY FURNISHED SEQUENCE LISTING

37. Any sequence listing which is not contained in the application as filed but which is furnished subsequently shall not go beyond the disclosure in the application as filed and shall be accompanied by a statement to that effect. This means that a sequence listing furnished subsequently to the filing of the application shall contain only those sequences that were disclosed in the application as filed.

38. Any sequence listing not contained in the application as filed does not form part of the application. However, the provisions of PCT Rules 13<sup>ter</sup>, 26.3 and 91 and PCT Article 34 would apply, so that it may be possible, subject to the applicable provisions, for a sequence listing contained in the application as filed to be corrected under PCT Rules 13<sup>ter</sup> or 26.3, rectified under PCT Rule 91 (in the case of an obvious error), or amended under PCT Article 34, or for a sequence listing to be submitted under PCT Article 34 as an amendment to the application.

#### COMPUTER READABLE FORM OF THE SEQUENCE LISTING

39. A copy of the sequence listing shall also be submitted in computer readable form, in addition to the sequence listing as contained in the application, whenever this is required by the competent Authority.

40. Any sequence listing in computer readable form submitted in addition to the written sequence listing shall be identical to the written sequence listing and shall be accompanied by a statement that "the information recorded in computer readable form is identical to the written sequence listing."

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 "texte libre du listage des séquences".

#### LISTAGE DES SÉQUENCES DÉPOSÉ ULTÉRIEUREMENT

37. 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.

38. Tout listage des séquences ne figurant pas dans la demande telle qu'elle a été déposée ne fait pas partie de la demande. Toutefois, les dispositions des règles 13<sup>er</sup>, 26.3 et 91 du règlement d'exécution du PCT et celles de l'article 34 du PCT s'appliquent normalement, de telle sorte qu'il est possible, sous réserve des dispositions applicables, qu'un listage des séquences figurant dans la demande telle qu'elle a été déposée fasse l'objet d'une correction en vertu de la règle 13<sup>er</sup> ou 26.3 du règlement d'exécution du PCT, d'une rectification en vertu de la règle 91 du règlement d'exécution du PCT (dans le cas d'une erreur évidente) ou d'une modification en vertu de l'article 34 du PCT, ou qu'un listage des séquences soit soumis en vertu de l'article 34 du PCT à titre de modification de la demande.

#### LISTAGE DES SÉQUENCES SOUS UNE FORME DÉCHIFFRABLE PAR ORDINATEUR

39. Outre le listage des séquences figurant dans la demande, une copie de ce même listage doit être fournie sous une forme déchiffable par ordinateur chaque fois que l'administration compétente l'exige.

40. Tout listage des séquences sous une forme déchiffable 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 déchiffable par ordinateur sont identiques à celles du listage des séquences présenté par écrit."

41. Das gesamte ausdrückbare Exemplar des Sequenzprotokolls muß in einer einzigen Datei enthalten sein, die nach Möglichkeit auf einer einzigen Diskette oder einem sonstigen von der zuständigen Behörde akzeptierten elektronischen Datenträger aufgezeichnet sein soll. Diese Datei ist mittels der IBM<sup>2</sup>-Codetabellen (Code Page) 437, 932<sup>3</sup> oder einer kompatiblen Codetabelle zu codieren. Eine Codetabelle, wie sie z. B. für japanische, chinesische, kyrillische, arabische, griechische oder hebräische Schriftzeichen benötigt wird, gilt als kompatibel, wenn sie das lateinische Alphabet und die arabischen Ziffern denselben Hexadezimalpositionen zuordnet wie die genannten Codetabellen.

42. Die maschinenlesbare Fassung soll mit spezieller Software wie PatentIn oder anderen kundenspezifischen Computerprogrammen erstellt werden; sie kann mit beliebigen Mitteln erstellt werden, solange das Sequenzprotokoll auf einer Diskette oder einem anderen von der zuständigen Behörde akzeptierten elektronischen Träger eingereicht wird und mit einem Betriebssystem für Personal Computer gelesen werden kann, das von der zuständigen Behörde akzeptiert wird.

43. Bei Verwendung von Disketten ist eine Datenkompression erlaubt, sofern die komprimierte Datei in einem selbstextrahierenden Format erstellt worden ist, das sich auf einem von der zuständigen Behörde akzeptierten Betriebssystem für Personal Computer selbst dekomprimiert.

44. Auf der Diskette oder einem anderen von der zuständigen Behörde akzeptierten elektronischen Träger ist ein Etikett fest anzubringen, auf dem von Hand oder mit der Maschine in Blockschrift der Name des Anmelders, die Bezeichnung der Erfindung, ein Aktenzeichen, der Zeitpunkt der Aufzeichnung der Daten, das Computerbetriebssystem und der Name der zuständigen Behörde eingetragen sind.

45. Wird die Diskette oder ein anderes von der zuständigen Behörde akzeptiertes elektronisches Medium erst nach dem Tag der Anmeldung

41. The entire printable copy of the sequence listing shall be contained within one electronic file preferably on a single diskette or any other electronic medium that is acceptable to the competent Authority. The file recorded on the diskette or any other electronic medium that is acceptable to the competent Authority shall be encoded using IBM<sup>2</sup> Code Page 437, IBM Code Page 932 or a compatible code page. A compatible code page, as would be required for, for example, Japanese, Chinese, Cyrillic, Arabic, Greek or Hebrew characters, is one that assigns the Roman alphabet and numerals to the same hexadecimal positions as do the specified code pages.

42. The computer readable form shall preferably be created by dedicated software such as PatentIn or other custom computer programs; it may be created by any means, as long as the sequence listing on a submitted diskette or any other electronic medium that is acceptable to the competent Authority is readable under a Personal Computer Operating system that is acceptable to the competent Authority.

43. File compression is acceptable when using diskette media, so long as the compressed file is in a self-extracting format that will decompress on a Personal Computer Operating system that is acceptable to the competent Authority.

44. The diskette or any other electronic medium that is acceptable to the competent Authority shall have a label permanently affixed thereto on which has been hand-printed, in block capitals or typed, the name of the applicant, the title of the invention, a reference number, the date on which the data were recorded, the computer operating system and the name of the competent Authority.

45. If the diskette or any other electronic medium that is acceptable to the competent Authority is submitted after the date of filing of an applica-

41. 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'administration compétente. 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 ou une page de code compatible. 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.

42. Le listage déchiffirable par ordinateur doit, de préférence, être créé par un logiciel spécialisé tel que PatentIn ou d'autres programmes informatiques personnalisés; il peut être créé par tout autre moyen dans la mesure où le listage des séquences figurant sur la disquette ou sur tout autre support électronique admis par l'administration compétente est déchiffirable sous un système d'exploitation d'ordinateur individuel lui aussi admis par l'administration compétente.

43. La compression d'un fichier est admise lorsque le support est une disquette, dans la mesure où le fichier compressé se présente sous un format auto-extractible qui se décompressera sur un système d'exploitation d'ordinateur personnel admis par l'administration compétente.

44. Sur la disquette ou sur tout autre support électronique admis par l'administration compétente doit être apposée 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, le système d'exploitation informatique et le nom de l'administration compétente.

45. Si la disquette ou tout autre support électronique admis par l'administration compétente est fournie après la date de dépôt de la

<sup>2</sup> IBM ist eine für die International Business Machine Corporation, Vereinigte Staaten von Amerika, eingetragene Handelsmarke.

<sup>3</sup> Die genannten Codetabellen gelten für Personal Computer praktisch als Standard.

<sup>2</sup> IBM is a registered trademark of International Business Machine Corporation, United States of America.

<sup>3</sup> The specified code pages are *de facto* standards for personal computers.

<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.

eingereicht, so sind auf den Etiketten auch der Anmeldetag und die Anmelde­nummer anzugeben.

46. Jeder Berichtigung des geschriebenen Sequenzprotokolls nach Regel 13<sup>ter</sup>.1 (a) i) oder 26.3 PCT, jeder Berichtigung eines offensichtlichen Fehlers im geschriebenen Sequenzprotokoll nach Regel 91 PCT und jeder nach Artikel 34 PCT vorgenommenen Änderung, die sich auf ein geschriebenes Sequenzprotokoll bezieht, ist ein maschinenlesbares Sequenzprotokoll beizufügen, in dem diese Berichtigung oder Änderung berücksichtigt worden ist.

tion, the labels shall also include the filing date of the application and the application number.

46. Any correction of the written sequence listing which is submitted under PCT Rules 13<sup>ter</sup>.1(a)(i) or 26.3, any rectification of an obvious error in the written sequence listing which is submitted under PCT Rule 91, or any amendment which includes a written sequence listing and which is submitted under PCT Article 34, shall be accompanied by a computer readable form of the sequence listing including any such correction, rectification or amendment.

demande, cette date et le numéro de la demande doivent aussi figurer sur les étiquettes.

46. Toute correction – en vertu des règles 13<sup>er</sup>.1.a)i) ou 26.3 du règlement d'exécution du PCT – du listage des séquences présenté par écrit, toute rectification – en vertu de la règle 91 du règlement d'exécution du PCT – d'une erreur évidente dans le listage des séquences présenté par écrit ou toute modification – en vertu de l'article 34 du PCT – qui comprend un listage des séquences présenté par écrit doit être accompagnée du listage des séquences sous une forme déchiffrable par ordinateur, comprenant la correction, la rectification ou la modification en question.

**Anhang 1****Numerische Kennzahlen**

In Sequenzprotokollen, die zu Anmeldungen eingereicht werden, dürfen nur die nachstehenden numerischen Kennzahlen verwendet werden. Die Überschriften der nachstehenden Datenelementrubriken dürfen in den Sequenzprotokollen nicht erscheinen.

Die numerischen Kennzahlen der obligatorischen Datenelemente, d. h. der Datenelemente, die in alle Sequenzprotokolle aufgenommen werden müssen (s. Nr. 25 dieses Standards: Kennziffern 110, 120, 160, 210, 211, 212, 213 und 400), und die numerischen Kennzahlen der Datenelemente, die in den in diesem Standard genannten Fällen aufgenommen werden müssen (s. Nr. 26,

27, 28, 29 und 30 dieses Standards: Kennzahlen 130, 140, 141, 150 und 151 sowie 220 bis 223), sind mit dem Buchstaben "O" gekennzeichnet.

Die numerischen Kennzahlen der fakultativen Datenelemente (s. Nr. 31 dieses Standards) sind mit dem Buchstaben "F" gekennzeichnet.

Numerische Kennzahl	Numerische Kennzahl Beschreibung	Obligatorisch (O) oder fakultativ (F)	Bemerkungen
<110>	Name des Anmelders	O	Ist der Name des Anmelders nicht in lateinischen Buchstaben geschrieben, so muß er – im Wege der Transliteration oder der Übersetzung ins Englische – auch in lateinischen Buchstaben angegeben werden.
<120>	Bezeichnung der Erfindung	O	
<130>	Aktenzeichen	O in den Fällen nach Nr. 26 Standard	Siehe Nr. 26 Standard.
<140>	Vorliegende Patentanmeldung	O in den Fällen nach Nr. 27 Standard	Siehe Nr. 27 Standard; die vorliegende Patentanmeldung ist zu kennzeichnen durch den zweibuchstabigen Code nach dem WIPO-Standard ST. 3, gefolgt von der Anmeldungsnummer (in dem Format, das von der Behörde für gewerblichen Rechtsschutz verwendet wird, bei der diese Patentanmeldung eingereicht wird) oder – bei internationalen Anmeldungen – von der internationalen Anmeldungsnummer.
<141>	Anmeldetag der vorliegenden Anmeldung	O in den Fällen nach Nr. 27 Standard	Siehe Nr. 27 Standard; das Datum ist entsprechend dem WIPO-Standard ST. 2 anzugeben (CCYY MM DD).
<150>	Frühere Patentanmeldung	O in den Fällen nach Nr. 28 Standard	Siehe Nr. 28 Standard; die frühere Patentanmeldung ist zu kennzeichnen durch den zweibuchstabigen Code entsprechend dem WIPO-Standard ST. 3, gefolgt von der Anmeldungsnummer (in dem Format, das von der Behörde für gewerblichen Rechtsschutz verwendet wird, bei der die frühere Patentanmeldung eingereicht wurde) oder – bei internationalen Anmeldungen – von der internationalen Anmeldungsnummer.
<151>	Anmeldetag der früheren Anmeldung	O in den Fällen nach Nr. 28 Standard	Siehe Nr. 28 Standard; das Datum ist entsprechend dem WIPO-Standard ST. 2 anzugeben (CCYY MM DD).
<160>	Anzahl der SEQ ID NOs	O	
<170>	Software	F	
<210>	Angaben zu SEQ ID NO: x	O	Anzugeben ist eine ganze Zahl, die die SEQ ID NO darstellt.
<211>	Länge	O	Sequenzlänge, ausgedrückt als Anzahl der Basenpaare oder Aminosäuren
<212>	Art	O	Art des in SEQ ID NO: x sequenzierten Moleküls, und zwar entweder DNA, RNA oder PRT; enthält eine Nucleotidsequenz

			sowohl DNA- als auch RNA-Fragmente, so ist "DNA" anzugeben; zusätzlich ist das kombinierte DNA-/RNA-Molekül im Merkmalsteil unter <220> bis <223> näher zu beschreiben.
<213>	Organismus	O	Gattung/Art (d. h. wissenschaftlicher Name), "künstliche Sequenz" oder "unbekannt"
<220>	Merkmal	O in den Fällen nach Nr. 29 und 30 Standard	Freilassen; siehe Nr. 29 und 30 Standard; Beschreibung biologisch signifikanter Stellen in der Sequenz gemäß SEQ ID NO: x) (kann je nach der Zahl der angegebenen Merkmale mehrmals vorkommen).
<221>	Name/Schlüssel	O in den Fällen nach Nr. 29 Standard	Siehe Nr. 29 Standard; es dürfen nur die in Anhang 2, Tabelle 5 oder 6 beschriebenen Schlüssel verwendet werden.
<222>	Lage	O in den Fällen nach Nr. 29 Standard	Siehe Nr. 29 Standard; <ul style="list-style-type: none"> <li>- von (Nummer der ersten Base/Aminosäure des Merkmals)</li> <li>- bis (Nummer der letzten Base/Aminosäure des Merkmals)</li> <li>- Basenpaare (Ziffern verweisen auf die Positionen der Basenpaare in einer Nucleotidsequenz)</li> <li>- Aminosäuren (Ziffern verweisen auf die Positionen der Aminosäurereste in einer Aminosäuresequenz)</li> <li>- Angabe, ob sich das Merkmal auf dem zum Strang des Sequenzprotokolls komplementären Strang befindet</li> </ul>
<223>	Sonstige Angaben	O in den Fällen nach Nr. 29 und 30 Standard	Siehe Nr. 29 und 30 Standard; sonstige relevante Angaben, wobei sprachneutrales Vokabular oder freier Text (möglichst in englischer Sprache) zu verwenden ist; freier Text ist im Hauptteil der Beschreibung in der dort verwendeten Sprache zu wiederholen (s. Nr. 36 Standard); enthält die Sequenz eine der in Anhang 2, Tabellen 2 und 4 aufgeführten modifizierten Basen oder modifizierten/seltenen L-Aminosäuren, so ist für diese Base oder Aminosäure das dazugehörige Symbol aus Anhang 2, Tabellen 2 und 4 zu verwenden.
<300>	Veröffentlichungsangaben	F	Freilassen; dieser Abschnitt ist für jede relevante Veröffentlichung zu wiederholen.
<301>	Verfasser	F	
<302>	Titel	F	Titel der Veröffentlichung
<303>	Zeitschrift	F	Name der Zeitschrift, in der die Daten veröffentlicht wurden
<304>	Band	F	Band der Zeitschrift, in dem die Daten veröffentlicht wurden
<305>	Heft	F	Nummer des Hefts der Zeitschrift, in dem die Daten veröffentlicht wurden
<306>	Seiten	F	Seiten der Zeitschrift, auf denen die Daten veröffentlicht wurden

<307>	Datum	F	Datum der Zeitschrift, an dem die Daten veröffentlicht wurden; Angabe nach Möglichkeit entsprechend dem WIPO-Standard ST. 2 (CCYY MM DD)
<308>	Eingangsnummer in der Datenbank	F	von der Datenbank zugeteilte Eingangsnummer einschließlich Datenbankbezeichnung
<309>	Datenbank-Eingabedatum	F	Datum der Eingabe in die Datenbank; Angabe entsprechend dem WIPO-Standard ST. 2 (CCYY MM DD)
<310>	Dokumentennummer	F	Nummer des Dokuments, nur bei Patentdokumenten; die vollständige Nummer hat nacheinander folgendes zu enthalten: den zweibuchstabigen Code entsprechend dem WIPO-Standard ST. 3, die Veröffentlichungsnummer entsprechend dem WIPO-Standard ST. 6 und den Code für die Dokumentenart nach dem WIPO-Standard ST. 16
<311>	Anmeldetag	F	Anmeldetag des Dokuments, nur bei Patentdokumenten; Angabe entsprechend WIPO-Standard ST. 2 (CCYY MM DD)
<312>	Veröffentlichungsdatum	F	Datum der Veröffentlichung des Dokuments; nur bei Patentdokumenten; Angabe entsprechend WIPO-Standard ST. 2 (CCYY MM DD)
<313>	Relevante Reste in SEQ ID NO: x von bis	F	
<400>	Sequenz	O	SEQ ID NO: x sollte in der der Sequenz vorausgehenden Zeile hinter der numerischen Kennzahl stehen (s. Anlage 3).

**Anhang 2****Symbole für Nucleotide und Aminosäuren und Merkmalstabellen**

Tabelle 1: Liste der Nucleotide

Symbol	Bedeutung	Ableitung der Bezeichnung
a	a	<u>A</u> denin
g	g	<u>G</u> uanin
c	c	<u>C</u> ytosin
t	t	<u>T</u> hymine
u	u	<u>U</u> racil
r	g oder a	Pu <u>r</u> in
y	t/u oder c	Pyrimid <u>i</u> n
m	a oder c	A <u>m</u> ino
k	g oder t/u	<u>K</u> eto
s	g oder c	<u>s</u> tarke Bindungen 3 H-Brücken
w	a oder t/u	schwache (e: <u>w</u> eak) Bindungen 2 H-Brücken
b	b oder c oder t/u	nicht a
d	a oder g oder t/u	nicht c
h	a oder c oder t/u	nicht g
v	a oder g oder c	nicht t, nicht u
n	a oder g oder c oder t/u, unbekannt oder sonstige	beliebig (e: <u>a</u> ny)



Tabelle 2: Liste der modifizierten Nucleotide

Symbol	Bedeutung
ac4c	4-Acetylcytidin
chm5u	5-(Carboxyhydroxymethyl)uridin
cm	2'-O-Methylcytidin
cmnm5s2u	5-Carboxymethylaminomethyl-2-thiouridin
cmnm5u	5-Carboxymethylaminomethyluridin
d	Dihydrouridin
fm	2'-O-Methylpseudouridin
gal q	beta, D-Galactosylqueuosin
gm	2'-O-Methylguanosin
i	Inosin
i6a	N6-Isopentenyladenosin
m1a	1-Methyladenosin
m1f	1-Methylpseudouridin
m1g	1-Methylguanosin
m1i	1-Methylinosin
m22g	2,2-Dimethylguanosin
m2a	2-Methyladenosin
m2g	2-Methylguanosin
m3c	3-Methylcytidin
m5c	5-Methylcytidin
m6a	N6-Methyladenosin
m7g	7-Methylguanosin
mam5u	5-Methylaminomethyluridin
mam5s2u	5-Methoxyaminomethyl-2-thiouridin
man q	beta, D-Mannosequeuosin
mcm5s2u	5-Methoxycarbonylmethyl-2-thiouridin
mcm5u	5-Methoxycarbonylmethyluridin
mo5u	5-Methoxyuridin
ms2i6a	2-Methylthio-N6-isopentenyladenosin
ms2t6a	N-((9-beta-D-Ribofuranosyl-2-methylthiopurin-6-yl) carbamoyl) threonin
mt6a	N-((9-beta-D-Ribofuranosylpurin-6-yl)N-methylcarbamoyl)threonin
mv	Uridin-5-oxyessigsäuremethylester
o5u	Uridin-5-oxyessigsäure (v)
osyw	Wybutoxosin
p	Pseudouridin

q	Qeuosin
s2c	2-Thiocytidin
s2t	5-Methyl-2-thiouridin
s2u	2-Thiouridin
s4u	4-Thiouridin
t	5-Methyluridin
t6a	N-((9-beta-D-Ribofuranosylpurin-6-yl) carbamoyl)threonin
tm	2'-O-Methyl-5-methyluridin
um	2'-O-Methyluridin
yw	Wybutosin
x	3-(3-Amino-3-carboxypropyl)uridin, (acp3)u

Tabelle 3: Liste der Aminosäuren

Symbol	Bedeutung
Ala	Alanin
Cys	Cystein
Asp	Asparaginsäure
Glu	Glutaminsäure
Phe	Phenylalanin
Gly	Glycin
His	Histidin
Ile	Isoleucin
Lys	Lysin
Leu	Leucin
Met	Methionin
Asn	Asparagin
Pro	Prolin
Gln	Glutamin
Arg	Arginin
Ser	Serin
Thr	Threonin
Val	Valin
Trp	Tryptophan
Tyr	Tyrosin
Asx	Asp oder Asn
Glx	Glu oder Gln
Xaa	unbekannt oder sonstige

Tabelle 4: Liste der modifizierten und seltenen Aminosäuren

Symbol	Bedeutung
Aad	2-Aminoadipinsäure
bAad	3-Aminoadipinsäure
bAla	beta-Alanin, beta-Aminopropionsäure
Abu	2-Aminobuttersäure
4Abu	4-Aminobuttersäure, Piperidinsäure
Acp	6-Aminocaprinsäure
Ahe	2-Aminoheptansäure
Aib	2-Aminoisobuttersäure
bAib	3-Aminoisobuttersäure
Apm	2-Aminopimelinsäure
Dbu	2,4-Diaminobuttersäure
Des	Desmosin
Dpm	2,2'-Diaminopimelinsäure
Dpr	2,3-Diaminopropionsäure
EtGly	N-Ethylglycin
EtAsn	N-Ethylasparagin
Hyl	Hydroxylysin
aHyl	allo-Hydroxylysin
3Hyp	3-Hydroxyprolin
4Hyp	4-Hydroxyprolin
Ide	Isodesmosin
alle	allo-Isoleucin
MeGly	N-Methylglycin, Sarkosin
Melle	N-Methylisoleucin
MeLys	6-N-Methyllysin
MeVal	N-Methylvalin
Nva	Norvalin
Nle	Norleucin
Orn	Ornithin

Tabelle 5: Liste der Merkmalschlüssel zu Nucleotidsequenzen

Schlüssel	Beschreibung
allele	Ein verwandtes Individuum oder ein verwandter Stamm enthält stabile alternative Formen desselben Gens und unterscheidet sich an dieser (und vielleicht an anderer) Stelle von der vorliegenden Sequenz.
attenuator	1. Region einer DNA, in der die Beendigung der Transkription reguliert wird und die Expression einiger bakterieller Operons gesteuert wird 2. zwischen dem Promotor und dem ersten Strukturgen liegender Sequenzabschnitt, der eine partielle Beendigung der Transkription bewirkt
C_region	konstante Region der leichten und schweren Immunglobulinketten und der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; enthält je nach Kette ein oder mehrere Exons
CAAT_signal	CAAT-Box; Teil einer konservierten Sequenz, der etwa 75 Basenpaare stromaufwärts vom Startpunkt der eukaryontischen Transkriptionseinheiten liegt und an der RNA-Polymerase-Bindung beteiligt sein kann; Konsensussequenz = GG (C oder T) CAATCT
CDS	codierende Sequenz; Sequenz von Nucleotiden, die mit der Sequenz der Aminosäuren in einem Protein übereinstimmt (beinhaltet Stopcodon); Merkmal schließt eine mögliche Translation der Aminosäure ein
conflict	Unabhängige Determinierungen "derselben" Sequenz unterscheiden sich an dieser Stelle oder in dieser Region voneinander.
D-loop	D-Schleife; Region innerhalb der mitochondrialen DNA, in der sich ein kürzeres RNA-Stück mit einem Strang der doppelsträngigen DNA paart und dabei den ursprünglichen Schwesterstrang in dieser Region verdrängt; dient auch zur Beschreibung der Verdrängung einer einzelsträngigen Region des DNA-Doppelstrangs durch einen einzelsträngigen Eindringling bei der durch ein recA-Protein ausgelösten Reaktion
D-segment	Diversity-Region der schweren Kette von Immunglobulin und der Beta-Kette eines T-Zell-Rezeptors
enhancer	eine als Cis-Element wirkende Sequenz, die die Aktivität (einiger) eukaryontischer Promotoren verstärkt und in beliebiger Richtung und Position zum Promotor (stromaufwärts oder -abwärts) funktioniert
exon	Region des Genoms, die für einen Teil der gespleißten mRNA codiert; kann 5'UTR, alle CDSs und 3'UTR enthalten
GC_signal	GC-Box; eine konservierte, GC-reiche Region stromaufwärts vom Startpunkt der eukaryontischen Transkriptionseinheiten, die in mehreren Kopien und in beiden Richtungen vorkommen kann; Konsensussequenz = GGGCGG
gene	biologisch signifikante Region, die als Gen bezeichnet wird und einen Namen trägt
iDNA	intervenierende DNA; DNA, die durch verschiedene Arten der Rekombination eliminiert wird
intron	DNA-Abschnitt, der transkribiert, aber beim Zusammenspleißen der ihn umrahmenden Sequenzen (Exons) aus dem Transkript wieder herausgeschnitten wird
J_segment	J-Kette (Verbindungskette) zwischen den leichten und den schweren Immunglobulinketten und den Alpha-, Beta- und Gamma-Ketten der T-Zell-Rezeptoren
LTR	lange Terminalwiederholung, eine an den beiden Enden einer gegebenen Sequenz direkt wiederholte Sequenz, wie sie für Retroviren typisch ist
mat_peptide	für ein reifes Peptid oder Protein codierende Sequenz; Sequenz, die im Anschluß an eine posttranslationale Modifizierung für das reife oder endgültige Peptid- oder Proteinprodukt codiert; schließt im Gegensatz zur entsprechenden CDS das Stopcodon nicht ein
misc_binding	Stelle in einer Nucleinsäure, die einen anderen Teil, der nicht durch einen anderen Bindungsschlüssel (primer_bind oder protein_bind) beschrieben werden kann, kovalent oder nicht kovalent bindet
misc_difference	Die Merkmalsequenz unterscheidet sich von der im Eintrag und kann nicht durch einen anderen Unterscheidungsschlüssel (conflict, unsure, old_sequence, mutation, variation, allele bzw. modified_base) beschrieben werden.

misc_feature	biologisch signifikante Region, die nicht durch einen anderen Merkmalschlüssel beschrieben werden kann; neues oder seltenes Merkmal
misc_recomb	Stelle, an der ein allgemeiner, ortsspezifischer oder replikativer Rekombinationsvorgang stattfindet, bei dem die DNA-Doppelhelix aufgebrochen und wieder zusammengefügt wird, und die nicht durch andere Rekombinationsschlüssel (iDNA oder Virion) oder den betreffenden Herkunftsschlüssel (/insertions_seq, /transposon, /proviral) beschrieben werden kann
misc_RNA	Transkript oder RNA-Produkt, das nicht durch andere RNA-Schlüssel (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 oder snRNA) beschrieben werden kann
misc_signal	Region, die ein Signal enthält, das die Genfunktion oder -expression steuert oder ändert, und die nicht durch andere Signalschlüssel (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator oder rep_origin) beschrieben werden kann
misc_structure	Sekundär-, Tertiär- oder sonstige Struktur oder Konformation, die nicht durch andere Strukturschlüssel (stem_loop oder D-loop) beschrieben werden kann
modified_base	Das angegebene Nucleotid ist ein modifiziertes Nucleotid und ist durch das angegebene Molekül (ausgedrückt durch die modifizierte Base) zu ersetzen.
mRNA	messenger-RNA (Boten-RNA); enthält eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon) und eine 3'-nichttranslatierte Region (3'UTR)
mutation	Ein verwandter Stamm weist an dieser Stelle eine plötzliche, erbliche Sequenzveränderung auf.
N_region	zusätzliche Nucleotide, die zwischen neu geordnete Immunglobulinabschnitte eingefügt werden
old_sequence	Die vorliegende Sequenz stellt eine geänderte Version der früher an dieser Stelle befindlichen Sequenz dar.
polyA_signal	Erkennungsregion, die zur Endonuclease-Spaltung eines RNA-Transkripts mit anschließender Polyadenylierung nötig ist; Konsensussequenz: AATAAA
polyA_site	Stelle auf einem RNA-Transkript, an der durch posttranslationale Polyadenylierung Adeninreste eingefügt werden
precursor_RNA	noch nicht gereifte RNA-Spezies; kann eine am 5'-Ende abzuschneidende Region (5'Clip), eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron), eine 3'-nichttranslatierte Region (3'UTR) und eine am 3'-Ende abzuschneidende Region (3'Clip) enthalten
prim_transcript	primäres (ursprüngliches, nicht prozessiertes) Transkript; enthält eine am 5'-Ende abzuschneidende Region (5'Clip), eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron), eine 3'-nichttranslatierte Region (3'UTR) und eine am 3'-Ende abzuschneidende Region (3'Clip)
primer_bind	nichtkovalente Primer-Bindungsstelle für die Initiierung der Replikation, Transkription oder reversen Transkription; enthält eine oder mehrere Stellen für synthetische Elemente, z. B. PCR-Primerelemente
promoter	Region auf einem DNA-Molekül, die an der Bindung der RNA-Polymerase beteiligt ist, die die Transkription initiiert
protein_bind	nichtkovalente Protein-Bindungsstelle auf der Nucleinsäure
RBS	ribosomale Bindungsstelle
repeat_region	Genomregion mit repetitiven Einheiten
repeat_unit	einzelnes Repeat (repetitive Einheit)
rep_origin	Replikationsursprung; Startpunkt der Duplikation der Nucleinsäure, durch die zwei identische Kopien entstehen
rRNA	reife ribosomale RNA; RNA-Komplex des Ribonucleoprotein-Partikels (Ribosom), der Aminosäuren zu Proteinen zusammenfügt
S_region	Switch-Region der schweren Immunglobulinketten; beteiligt am Umbau schwerer DNA-Ketten, der zur Expression einer anderen Immunglobulin-Klasse aus derselben B-Zelle führt

satellite	viele tandemartig hintereinandergeschaltete (identische oder verwandte) Repeats einer kurzen basischen repetitiven Einheit; viele davon unterscheiden sich in der Basenzusammensetzung oder einer anderen Eigenschaft vom Genomdurchschnitt und können so von der Hauptmasse der genomischen DNA abgetrennt werden
scRNA	kleine cytoplasmische RNA; eines von mehreren kleinen cytoplasmischen RNA-Molekülen im Cytoplasma und (manchmal) im Zellkern eines Eukaryonten
sig_peptide	für ein Signalpeptid codierende Sequenz; Sequenz, die für eine N-terminale Domäne eines sekretorischen Proteins codiert; diese Domäne spielt bei der Anheftung des naszierenden Polypeptids an die Membran eine Rolle; Leader-Sequenz
snRNA	kleine Kern-RNA; eine der vielen kleinen RNA-Formen, die nur im Zellkern vorkommen; einige der snRNAs spielen beim Spleißen oder bei anderen RNA-verarbeitenden Reaktionen eine Rolle
source	Bezeichnet die biologische Herkunft des genannten Sequenzabschnitts; die Angabe dieses Schlüssels ist obligatorisch; jeder Eintrag muß mindestens einen Herkunftsschlüssel aufweisen, der die gesamte Sequenz umfaßt; es dürfen zu jeder Sequenz auch mehrere Herkunftsschlüssel angegeben werden.
stem_loop	Haarnadelschleife; eine Doppelhelix-Region, die durch Basenpaarung zwischen benachbarten (invertierten) komplementären Sequenzen in einem RNA- oder DNA-Einzelstrang entsteht
STS	Sequence Tagged Site; kurze, nur als Einzelkopie vorkommende DNA-Sequenz, die einen Kartierungspunkt auf dem Genom bezeichnet und durch PCR ermittelt werden kann; eine Region auf dem Genom kann durch Bestimmung der Reihenfolge der STSs kartiert werden
TATA_signal	TATA-Box; Goldberg-Hogness-Box; ein konserviertes AT-reiches Septamer, das sich rund 25 Basenpaare vor dem Startpunkt jeder eukaryontischen RNA-Polymerase-II-Transkriptionseinheit befindet und bei der Positionierung des Enzyms für eine korrekte Initiation eine Rolle spielen kann; Konsensussequenz = TATA(A oder T)A(A oder T)
terminator	DNA-Sequenz, die entweder am Ende des Transkripts oder neben einer Promotor-Region liegt und bewirkt, daß die RNA-Polymerase die Transkription beendet; kann auch die Bindungsstelle eines Repressor-Proteins sein
transit_peptide	für Transitpeptid codierende Sequenz; Sequenz, die für eine N-terminale Domäne eines im Zellkern codierten Organellen-Proteins codiert; diese Domäne ist an der posttranslationalen Einschleusung des Proteins in die Organelle beteiligt
tRNA	reife transfer-RNA, ein kleines RNA-Molekül (75–85 Basen lang), das die Translation einer Nucleinsäure-Sequenz in eine Aminosäure-Sequenz vermittelt
unsure	Der Autor kennt die Sequenz in dieser Region nicht genau.
V_region	variable Region der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; codiert für das variable Aminoende; kann aus V-, D-, N- und J-Abschnitten bestehen
V-segment	variabler Abschnitt der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; codiert für den Großteil der variablen Region (V_region) und die letzten Aminosäuren des Leader-Peptids
variation	Ein verwandter Stamm enthält stabile Mutationen desselben Gens (z. B. RFLPs, Polymorphismen usw.), die sich an dieser (und möglicherweise auch an anderer) Stelle von der vorliegenden Sequenz unterscheiden.
3'clip	3'-äußerste Region eines Precursor-Transkripts, die bei der Prozessierung abgeschnitten wird
3'UTR	Region am 3'-Ende eines reifen Transkripts (nach dem Stopcodon), die nicht in ein Protein translatiert wird
5'clip	5'-äußerste Region eines Precursor-Transkripts, die bei der Prozessierung abgeschnitten wird
5'UTR	Region am 5'-Ende eines reifen Transkripts (vor dem Initiationscodon), die nicht in ein Protein translatiert wird
-10_signal	Pribnow-Box; konservierte Region rund 10 Basenpaare stromaufwärts vom Startpunkt der bakteriellen Transkriptionseinheiten, die bei der Bindung der RNA-Polymerase eine Rolle spielt; Konsensussequenz = TAtAAT
-35_signal	konserviertes Hexamer rund 35 Basenpaare stromaufwärts vom Startpunkt der bakteriellen Transkriptionseinheiten; Konsensussequenz = TTGACa [ ] oder TGTTGACA [ ]

Tabelle 6: Liste der Merkmalschlüssel zu Proteinsequenzen

Schlüssel	Beschreibung
CONFLICT	In den einzelnen Unterlagen ist von verschiedenen Sequenzen die Rede.
VARIANT	Den Angaben der Autoren zufolge gibt es Sequenzvarianten.
VARSPLIC	Beschreibung von Sequenzvarianten, die durch alternatives Spleißen entstanden sind
MUTAGEN	experimentell veränderte Stelle
MOD_RES	posttranslationale Modifikation eines Rests
ACETYLATION	N-terminale oder sonstige
AMIDATION	in der Regel am C-Terminus eines reifen aktiven Peptids
BLOCKED	unbestimmte Gruppe, die das N- oder C-terminale Ende blockiert
FORMYLATION	des N-terminalen Methionin
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	von Asparagin, Asparaginsäure, Prolin oder Lysin
METHYLATION	in der Regel von Lysin oder Arginin
PHOSORYLATION	von Serin, Threonin, Tyrosin, Asparaginsäure oder Histidin
PYRROLIDONE	N-terminales Glutamat, das ein internes cyclisches Lactam gebildet hat
CARBOXYLIC ACID SULFATATION	in der Regel von Tyrosin
LIPID	kovalente Bindung einer Lipidhälfte
MYRISTATE	Myristat-Gruppe, die durch eine Amidbindung an den N-terminalen Glycin-Rest der reifen Form eines Proteins oder an einen internen Lysin-Rest gebunden ist
PALMITATE	Palmitat-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest oder durch eine Esterbindung an einen Serin- oder Threonin-Rest gebunden ist
FARNESYL	Farnesyl-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest gebunden ist
GERANYL-GERANYL	Geranyl-geranyl-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest gebunden ist
GPI-ANCHOR	Glykosyl-phosphatidylinositol-(GPI-)Gruppe, die an die alpha-Carboxylgruppe des C-terminalen Rests der reifen Form eines Proteins gebunden ist
N-ACYL DIGLYCERIDE	N-terminales Cystein der reifen Form eines prokaryontischen Lipoproteins mit einer amidgebundenen Fettsäure und einer Glyceryl-Gruppe, an die durch Esterbindungen zwei Fettsäuren gebunden sind
DISULFID	Disulfidbindung; den "VON"- und den "BIS"-Endpunkt bilden die beiden Reste, die durch eine ketteninterne Disulfidbindung verbunden sind; sind der "VON"- und der "BIS"-Endpunkt identisch, liegt die Disulfidbindung innerhalb der Kette, und die Art der Quervernetzung ist im Beschreibungsfeld anzugeben
THIOLEST	Thiolesterbindung; den "VON"- und den "BIS"-Endpunkt bilden die beiden Reste, die durch die Thiolesterbindung verbunden sind
THIOETH	Thioetherbindung; den "VON"- und den "BIS"-Endpunkt bilden die beiden Reste, die durch die Thioetherbindung verbunden sind
CARBOHYD	Glykosylierungs-Stelle; die Art des Kohlenhydrats (sofern bekannt) ist im Beschreibungsfeld anzugeben
METAL	Bindungsstelle für ein Metallion; die Art des Metalls ist im Beschreibungsfeld anzugeben

BINDING	Bindungsstelle für eine beliebige chemische Gruppe (Coenzym, prosthetische Gruppe usw.); die Art der Gruppe ist im Beschreibungsfeld anzugeben
SIGNAL	Bereich einer Signalsequenz (Präpeptid)
TRANSIT	Bereich eines Transit-Peptids (mitochondriales, chloroplastisches oder für Microbodies)
PROPEP	Bereich eines Propeptids
CHAIN	Bereich einer Polypeptid-Kette im reifen Protein
PEPTIDE	Bereich eines freigesetzten aktiven Peptids
DOMAIN	Bereich einer wichtigen Domäne auf der Sequenz; die Art dieser Domäne ist im Beschreibungsfeld anzugeben
CA_BIND	Bereich einer Calcium-bindenden Region
DNA_BIND	Bereich einer DNA-bindenden Region
NP_BIND	Bereich einer Nucleotidphosphat-bindenden Region; die Art des Nucleotidphosphats ist im Beschreibungsfeld anzugeben
TRANSMEM	Bereich einer Transmembran-Region
ZN_FING	Bereich einer Zink-Finger-Region
SIMILAR	Grad der Ähnlichkeit mit einer anderen Proteinsequenz; im Beschreibungsfeld sind genaue Angaben über diese Sequenz zu machen
REPEAT	Bereich einer internen Sequenzwiederholung
HELIX	Sekundärstruktur: Helices, z. B. Alpha-Helix, 3(10)-Helix oder Pi-Helix
STRAND	Sekundärstruktur: Beta-Strang, z. B. durch Wasserstoff-Brückenbindungen stabilisierter Beta-Strang, oder Rest in einer isolierten Beta-Brücke
TURN	Sekundärstruktur: Schleife, z. B. durch Wasserstoff-Brückenbindungen stabilisierte Schleife (3-, 4- oder 5-Schleife)
ACT_SITE	Aminosäure(n), die bei der Aktivität eines Enzyms mitwirkt (mitwirken)
SITE	irgendeine andere wichtige Stelle auf der Sequenz
INIT_MET	die Sequenz beginnt bekanntermaßen mit einem Start-Methionin
NON_TER	Der Rest am Sequenzanfang oder -ende ist nicht der Terminalrest; steht er an der Position 1, so bedeutet das, daß diese nicht der N-Terminus des vollständigen Moleküls ist; steht er an letzter Position, so ist diese Position nicht der C-Terminus des vollständigen Moleküls; für diesen Schlüssel gibt es kein Beschreibungsfeld.
NON_CONS	nicht aufeinanderfolgende Reste; zeigt an, daß zwei Reste in einer Sequenz nicht aufeinanderfolgen, sondern daß zwischen ihnen einige nichtsequenzierte Reste liegen
UNSURE	Unsicherheiten in der Sequenz; mit diesem Schlüssel werden Regionen einer Sequenz beschrieben, bei der sich der Autor bezüglich der Sequenzzuweisung nicht sicher ist



### Anhang 3

#### Muster eines Sequenzprotokolls

```

<110>      Smith, John; Smithgene Inc.
<120>      Beispiel für ein Sequenzprotokoll
<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  cgcgctcctc  tcgcgctctc  ctctcgctct  cctctcgctc          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
    
```

<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> Künstliche Sequenz

<220>  
 <223> Peptid, das nach Größe und Polarität als Verbindung zwischen den Alpha- und Beta-Ketten des Proteins XYZ konzipiert ist

<400> 3

Met	Val	Asn	Leu	Glu	Pro	Met	His	Thr	Glu	Ile
1				5					10	

<210> 4  
 <400> 4  
 000

## Appendix 1

### Numeric identifiers

Only numeric identifiers as defined below may be used in sequence listings submitted in applications. The text of the data element headings given below shall not be included in the sequence listings.

Numeric identifiers of mandatory data elements, that is, data elements which must be included in all sequence listings (see paragraph 25 of this Standard: items 110, 120, 160, 210, 211, 212, 213 and 400) and numeric identifiers of data elements which must be included in circumstances specified in this Standard (see paragraphs 26, 27, 28, 29 and 30

of this Standard: items 130, 140, 141, 150 and 151, and 220 to 223) are marked by the symbol "M".

Numeric identifiers of optional data elements (see paragraph 31 of this Standard) are marked by the symbol "O".

Numeric Identifier	Numeric Identifier Description	Mandatory (M) or Optional (O)	Comment
<110>	Applicant name	M	where the name of the applicant is written in characters other than those of the Latin alphabet, the same shall also be indicated in characters of the Latin alphabet either as a mere transliteration or through translation into English
<120>	Title of invention	M	
<130>	File reference	M, in the circumstances specified in paragraph 26 of this Standard	see paragraph 26 of this Standard
<140>	Current patent application	M, in the circumstances specified in paragraph 27 of this Standard	see paragraph 27 of this Standard; the current patent application shall be identified, in the following order, by the two-letter code indicated in accordance with WIPO Standard ST.3 and the application number (in the format used by the industrial property Office with which the current patent application is filed) or, for an international application, by the international application number
<141>	Current filing date	M, in the circumstances specified in paragraph 27 of this Standard	see paragraph 27 of this Standard; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<150>	Earlier patent application	M, in the circumstances specified in paragraph 28 of this Standard	see paragraph 28 of this Standard; the earlier patent application shall be identified, in the following order, by the two-letter code indicated in accordance with WIPO Standard ST.3 and the application number (in the format used by the industrial property Office with which the earlier patent application was filed) or, for an international application, by the international application number
<151>	Earlier application filing date	M, in the circumstances specified in paragraph 28 of this Standard	see paragraph 28 of this Standard; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<160>	Number of SEQ ID NOs	M	
<170>	Software	O	
<210>	Information for SEQ ID NO: x	M	response shall be an integer representing the SEQ ID NO shown
<211>	Length	M	sequence length expressed in number of base pairs or amino acids
<212>	Type	M	type of molecule sequenced in SEQ ID NO: x, either DNA, RNA or PRT; if a nucleotide sequence contains both DNA

			and RNA fragments, the value shall be "DNA"; in addition, the combined DNA/RNA molecule shall be further described in the <220> to <223> feature section
<213>	Organism	M	Genus Species (that is, scientific name) or "Artificial Sequence" or "Unknown"
<220>	Feature	M, in the circumstances specified in paragraph 29 and 30 of this Standard	leave blank; see paragraphs 29 and 30 of this Standard; description of points of biological significance in the sequence in SEQ ID NO: x) (may be repeated depending on the number of features indicated)
<221>	Name/key	M, in the circumstances specified in paragraph 29 of this Standard	see paragraph 29 of this Standard; only those keys as described in Table 5 or 6 of Appendix 2 shall be used
<222>	Location	M, in the circumstances specified in paragraph 29 of this Standard	see paragraph 29 of this Standard; <ul style="list-style-type: none"> <li>- from (number of first base/amino acid in the feature)</li> <li>- to (number of last base/amino acid in the feature)</li> <li>- base pairs (numbers refer to positions of base pairs in a nucleotide sequence)</li> <li>- amino acids (numbers refer to positions of amino acid residues in an amino acid sequence)</li> <li>- whether feature is located on the complementary strand to that filed in the sequence listing</li> </ul>
<223>	Other information	M, in the circumstances specified in paragraphs 29 and 30 of this Standard	see paragraphs 29 and 30 of this Standard; any other relevant information, using language neutral vocabulary, or free text (preferably in English); any free text is to be repeated in the main part of the description in the language thereof (see paragraph 36 of this Standard); where any modified base or modified/unusual L-amino acid appearing in Appendix 2, Tables 2 and 4, is in the sequence, the symbol associated with that base or amino acid from Appendix 2, Tables 2 and 4, should be used
<300>	Publication information	O	leave blank; repeat section for each relevant publication
<301>	Authors	O	
<302>	Title	O	title of publication
<303>	Journal	O	journal name in which data published
<304>	Volume	O	journal volume in which data published
<305>	Issue	O	journal issue number in which data published
<306>	Pages	O	journal page numbers on which data published
<307>	Date	O	journal date on which data published; if possible, the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<308>	Database accession number	O	accession number assigned by database including database name

<309>	Database entry date	O	date of entry in database; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<310>	Document number	O	document number, for patent type citations only; the full document shall specify, in the following order, the two-letter code indicated in accordance with WIPO Standard ST.3, the publication number indicated in accordance with WIPO Standard ST.6, and the kind-of-document code indicated in accordance with WIPO Standard ST.16
<311>	Filing date	O	document filing date, for patent-type citations only; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<312>	Publication date	O	document publication date; for patent-type citations only; the date shall be indicated in accordance with WIPO Standard ST.2 (CCYY MM DD)
<313>	Relevant residues in SEQ ID NO: x: from to	O	
<400>	Sequence	M	SEQ ID NO: x should follow the numeric identifier and should appear on the line preceding the sequence (see Appendix 3)

## Appendix 2

### Nucleotide and amino acid symbols and feature table

Table 1: List of Nucleotides

Symbol	Meaning	Origin of designation
a	a	<u>a</u> denine
g	g	g <u>u</u> anine
c	c	<u>c</u> ytosine
t	t	<u>t</u> hymine
u	u	<u>u</u> racil
r	g or a	pu <u>r</u> ine
y	t/u or c	py <u>r</u> imidine
m	a or c	<u>a</u> mino
k	g or t/u	<u>k</u> eto
s	g or c	<u>s</u> trong interactions 3H-bonds
w	a or t/u	<u>w</u> weak interactions 2H-bonds
b	g or c or t/u	not a
d	a or g or t/u	not c
h	a or c or t/u	not g
v	a or g or c	not t, not u
n	a or g or c or t/u, unknown, or other	<u>a</u> ny

Table 2: List of Modified Nucleotides

Symbol	Meaning
ac4c	4-acetylcytidine
chm5u	5-(carboxyhydroxymethyl)uridine
cm	2'-O-methylcytidine
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine
cmnm5u	5-carboxymethylaminomethyluridine
d	dihydrouridine
fm	2'-O-methylpseudouridine
gal q	beta, D-galactosylqueuosine
gm	2'-O-methylguanosine
i	inosine
i6a	N6-isopentenyladenosine
m1a	1-methyladenosine
m1f	1-methylpseudouridine
m1g	1-methylguanosine
m1i	1-methylinosine
m22g	2,2-dimethylguanosine
m2a	2-methyladenosine
m2g	2-methylguanosine
m3c	3-methylcytidine
m5c	5-methylcytidine
m6a	N6-methyladenosine
m7g	7-methylguanosine
mam5u	5-methylaminomethyluridine
mam5s2u	5-methoxyaminomethyl-2-thiouridine
man q	beta, D-mannosylqueuosine
mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
mcm5u	5-methoxycarbonylmethyluridine
mo5u	5-methoxyuridine
ms2i6a	2-methylthio-N6-isopentenyladenosine
ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methylcarbamoyl)threonine
mv	uridine-5-oxyacetic acid-methylester
o5u	uridine-5-oxyacetic acid
osyw	wybutoxosine
p	pseudouridine

q	queuosine
s2c	2-thiocytidine
s2t	5-methyl-2-thiouridine
s2u	2-thiouridine
s4u	4-thiouridine
t	5-methyluridine
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)-carbamoyl)threonine
tm	2'-O-methyl-5-methyluridine
um	2'-O-methyluridine
yw	wybutosine
x	3-(3-amino-3-carboxy-propyl)uridine, (acp3)u

Table 3: List of Amino Acids

Symbol	Meaning
Ala	Alanine
Cys	Cysteine
Asp	Aspartic Acid
Glu	Glutamic Acid
Phe	Phenylalanine
Gly	Glycine
His	Histidine
Ile	Isoleucine
Lys	Lysine
Leu	Leucine
Met	Methionine
Asn	Asparagine
Pro	Proline
Gln	Glutamine
Arg	Arginine
Ser	Serine
Thr	Threonine
Val	Valine
Trp	Tryptophan
Tyr	Tyrosine
Asx	Asp or Asn
Glx	Glu or Gln
Xaa	unknown or other



**Table 4: List of Modified and Unusual Amino Acids**

Symbol	Meaning
Aad	2-Aminoadipic acid
bAad	3-Aminoadipic acid
bAla	beta-Alanine, beta-Aminopropionic acid
Abu	2-Aminobutyric acid
4Abu	4-Aminobutyric acid, piperidinic acid
Acp	6-Aminocaproic acid
Ahe	2-Aminoheptanoic acid
Aib	2-Aminoisobutyric acid
bAib	3-Aminoisobutyric acid
Apm	2-Aminopimelic acid
Dbu	2,4-Diaminobutyric acid
Des	Desmosine
Dpm	2,2'-Diaminopimelic acid
Dpr	2,3-Diaminopropionic acid
EtGly	N-Ethylglycine
EtAsn	N-Ethylasparagine
Hyl	Hydroxylysine
aHyl	allo-Hydroxylysine
3Hyp	3-Hydroxyproline
4Hyp	4-Hydroxyproline
Ide	Isodesmosine
alle	allo-Isoleucine
MeGly	N-Methylglycine, sarcosine
Melle	N-Methylisoleucine
MeLys	6-N-Methyllysine
MeVal	N-Methylvaline
Nva	Norvaline
Nle	Norleucine
Orn	Ornithine

**Table 5: List of Feature Keys Related to Nucleotide Sequences**

Key	Description
allele	a related individual or strain contains stable, alternative forms of the same gene which differs from the presented sequence at this location (and perhaps others)
attenuator	(1) region of DNA at which regulation of termination of transcription occurs, which controls the expression of some bacterial operons; (2) sequence segment located between the promoter and the first structural gene that causes partial termination of transcription
C_region	constant region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; includes one or more exons depending on the particular chain
CAAT_signal	CAAT box; part of a conserved sequence located about 75 bp up-stream of the start point of eukaryotic transcription units which may be involved in RNA polymerase binding; consensus=GG (C or T) CAATCT
CDS	coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon); feature includes amino acid conceptual translation
conflict	independent determinations of the "same" sequence differ at this site or region
D-loop	displacement loop; a region within mitochondrial DNA in which a short stretch of RNA is paired with one strand of DNA, displacing the original partner DNA strand in this region; also used to describe the displacement of a region of one strand of duplex DNA by a single stranded invader in the reaction catalyzed by RecA protein
D-segment	diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain
enhancer	a cis-acting sequence that increases the utilization of (some) eukaryotic promoters, and can function in either orientation and in any location (upstream or downstream) relative to the promoter
exon	region of genome that codes for portion of spliced mRNA; may contain 5'UTR, all CDSs, and 3'UTR
GC_signal	GC box; a conserved GC-rich region located upstream of the start point of eukaryotic transcription units which may occur in multiple copies or in either orientation; consensus=GGGCGG
gene	region of biological interest identified as a gene and for which a name has been assigned
iDNA	intervening DNA; DNA which is eliminated through any of several kinds of recombination
intron	a segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it
J_segment	joining segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains
LTR	long terminal repeat, a sequence directly repeated at both ends of a defined sequence, of the sort typically found in retroviruses
mat_peptide	mature peptide or protein coding sequence; coding sequence for the mature or final peptide or protein product following post-translational modification; the location does not include the stop codon (unlike the corresponding CDS)
misc_binding	site in nucleic acid which covalently or non-covalently binds another moiety that cannot be described by any other Binding key (primer_bind or protein_bind)
misc_difference	feature sequence is different from that presented in the entry and cannot be described by any other Difference key (conflict, unsure, old_sequence, mutation, variation, allele, or modified_base)
misc_feature	region of biological interest which cannot be described by any other feature key; a new or rare feature
misc_recomb	site of any generalized, site-specific or replicative recombination event where there is a breakage and reunion of duplex DNA that cannot be described by other recombination keys (iDNA and virion) or qualifiers of source key (/insertion_seq, /transposon, /proviral)

misc_RNA	any transcript or RNA product that cannot be defined by other RNA keys (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, and snRNA)
misc_signal	any region containing a signal controlling or altering gene function or expression that cannot be described by other Signal keys (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator, and rep_origin)
misc_structure	any secondary or tertiary structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop)
modified_base	the indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value)
mRNA	messenger RNA; includes 5' untranslated region (5'UTR), coding sequences (CDS, exon) and 3' untranslated region (3'UTR)
mutation	a related strain has an abrupt, inheritable change in the sequence at this location
N_region	extra nucleotides inserted between rearranged immunoglobulin segments
old_sequence	the presented sequence revises a previous version of the sequence at this location
polyA_signal	recognition region necessary for endonuclease cleavage of an RNA transcript that is followed by polyadenylation; consensus=AATAAA
polyA_site	site on an RNA transcript to which will be added adenine residues by post-transcriptional polyadenylation
precursor_RNA	any RNA species that is not yet the mature RNA product; may include 5' clipped region (5'clip), 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron), 3' untranslated region (3'UTR), and 3' clipped region (3'clip)
prim_transcript	primary (initial, unprocessed) transcript; includes 5' clipped region (5'clip), 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron), 3' untranslated region (3'UTR), and 3' clipped region (3'clip)
primer_bind	non-covalent primer binding site for initiation of replication, transcription, or reverse transcription; includes site(s) for synthetic, for example, PCR primer elements
promoter	region on a DNA molecule involved in RNA polymerase binding to initiate transcription
protein_bind	non-covalent protein binding site on nucleic acid
RBS	ribosome binding site
repeat_region	region of genome containing repeating units
repeat_unit	single repeat element
rep_origin	origin of replication; starting site for duplication of nucleic acid to give two identical copies
rRNA	mature ribosomal RNA; the RNA component of the ribonucleoprotein particle (ribosome) which assembles amino acids into proteins
S_region	switch region of immunoglobulin heavy chains; involved in the rearrangement of heavy chain DNA leading to the expression of a different immunoglobulin class from the same B-cell
satellite	many tandem repeats (identical or related) of a short basic repeating unit; many have a base composition or other property different from the genome average that allows them to be separated from the bulk (main band) genomic DNA
scRNA	small cytoplasmic RNA; any one of several small cytoplasmic RNA molecules present in the cytoplasm and (sometimes) nucleus of a eukaryote
sig_peptide	signal peptide coding sequence; coding sequence for an N-terminal domain of a secreted protein; this domain is involved in attaching nascent polypeptide to the membrane; leader sequence
snRNA	small nuclear RNA; any one of many small RNA species confined to the nucleus; several of the snRNAs are involved in splicing or other RNA processing reactions

source	identifies the biological source of the specified span of the sequence; this key is mandatory; every entry will have, as a minimum, a single source key spanning the entire sequence; more than one source key per sequence is permissible
stem_loop	hairpin; a double-helical region formed by base-pairing between adjacent (inverted) complementary sequences in a single strand of RNA or DNA
STS	Sequence Tagged Site; short, single-copy DNA sequence that characterizes a mapping landmark on the genome and can be detected by PCR; a region of the genome can be mapped by determining the order of a series of STSs
TATA_signal	TATA box; Goldberg-Hogness box; a conserved AT-rich septamer found about 25 bp before the start point of each eukaryotic RNA polymerase II transcript unit which may be involved in positioning the enzyme for correct initiation; consensus=TATA(A or T)A(A or T)
terminator	sequence of DNA located either at the end of the transcript or adjacent to a promoter region that causes RNA polymerase to terminate transcription; may also be site of binding of repressor protein
transit_peptide	transit peptide coding sequence; coding sequence for an N-terminal domain of a nuclear-encoded organellar protein; this domain is involved in post-translational import of the protein into the organelle
tRNA	mature transfer RNA, a small RNA molecule (75–85 bases long) that mediates the translation of a nucleic acid sequence into an amino acid sequence
unsure	author is unsure of exact sequence in this region
V_region	variable region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for the variable amino terminal portion; can be made up from V_segments, D_segments, N_regions, and J_segments
V_segment	variable segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for most of the variable region (V_region) and the last few amino acids of the leader peptide
variation	a related strain contains stable mutations from the same gene (for example, RFLPs, polymorphisms, etc.) which differ from the presented sequence at this location (and possibly others)
3'clip	3'-most region of a precursor transcript that is clipped off during processing
3'UTR	region at the 3' end of a mature transcript (following the stop codon) that is not translated into a protein
5'clip	5'-most region of a precursor transcript that is clipped off during processing
5'UTR	region at the 5' end of a mature transcript (preceding the initiation codon) that is not translated into a protein
-10_signal	pribnow box; a conserved region about 10 bp upstream of the start point of bacterial transcription units which may be involved in binding RNA polymerase; consensus=TAtAaT
-35_signal	a conserved hexamer about 35 bp upstream of the start point of bacterial transcription units; consensus=TTGACa [ ] or TGTTGACA [ ]

**Table 6: List of Feature Keys Related to Protein Sequences**

Key	Description
CONFLICT	different papers report differing sequences
VARIANT	authors report that sequence variants exist
VARSP LIC	description of sequence variants produced by alternative splicing
MUTAGEN	site which has been experimentally altered
MOD_RES	post-translational modification of a residue
ACETYLATION	N-terminal or other
AMIDATION	generally at the C-terminal of a mature active peptide
BLOCKED	undetermined N- or C-terminal blocking group
FORMYLATION	of the N-terminal methionine
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	of asparagine, aspartic acid, proline or lysine
METHYLATION	generally of lysine or arginine
PHOSPHORYLATION	of serine, threonine, tyrosine, aspartic acid or histidine
PYRROLIDONE CARBOXYLIC ACID	N-terminal glutamate which has formed an internal cyclic lactam
SULFATATION	generally of tyrosine
LIPID	covalent binding of a lipidic moiety
MYRISTATE	myristate group attached through an amide bond to the N-terminal glycine residue of the mature form of a protein or to an internal lysine residue
PALMITATE	palmitate group attached through a thioether bond to a cysteine residue or through an ester bond to a serine or threonine residue
FARNESYL	farnesyl group attached through a thioether bond to a cysteine residue
GERANYL-GERANYL	geranyl-geranyl group attached through a thioether bond to a cysteine residue
GPI-ANCHOR	glycosyl-phosphatidylinositol (GPI) group linked to the alpha-carboxyl group of the C-terminal residue of the mature form of a protein
N-ACYL DIGLYCERIDE	N-terminal cysteine of the mature form of a prokaryotic lipoprotein with an amide-linked fatty acid and a glyceryl group to which two fatty acids are linked by ester linkages
DISULFID	disulfide bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by an intra-chain disulfide bond; if the 'FROM' and 'TO' endpoints are identical, the disulfide bond is an interchain one and the description field indicates the nature of the cross-link
THIOLEST	thiolester bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by the thiolester bond
THIOETH	thioether bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by the thioether bond
CARBOHYD	glycosylation site; the nature of the carbohydrate (if known) is given in the description field
METAL	binding site for a metal ion; the description field indicates the nature of the metal
BINDING	binding site for any chemical group (co-enzyme, prosthetic group, etc.); the chemical nature of the group is given in the description field

SIGNAL	extent of a signal sequence (prepeptide)
TRANSIT	extent of a transit peptide (mitochondrial, chloroplastic, or for a microbody)
PROPEP	extent of a propeptide
CHAIN	extent of a polypeptide chain in the mature protein
PEPTIDE	extent of a released active peptide
DOMAIN	extent of a domain of interest on the sequence; the nature of that domain is given in the description field
CA_BIND	extent of a calcium-binding region
DNA_BIND	extent of a DNA-binding region
NP_BIND	extent of a nucleotide phosphate binding region; the nature of the nucleotide phosphate is indicated in the description field
TRANSMEM	extent of a transmembrane region
ZN_FING	extent of a zinc finger region
SIMILAR	extent of a similarity with another protein sequence; precise information, relative to that sequence is given in the description field
REPEAT	extent of an internal sequence repetition
HELIX	secondary structure: Helices, for example, Alpha-helix, 3(10) helix, or Pi-helix
STRAND	secondary structure: Beta-strand, for example, Hydrogen bonded beta-strand, or Residue in an isolated beta-bridge
TURN	secondary structure Turns, for example, H-bonded turn (3-turn, 4-turn or 5-turn)
ACT_SITE	amino acid(s) involved in the activity of an enzyme
SITE	any other interesting site on the sequence
INIT_MET	the sequence is known to start with an initiator methionine
NON_TER	the residue at an extremity of the sequence is not the terminal residue; if applied to position 1, this signifies that the first position is not the N-terminus of the complete molecule; if applied to the last position, it signifies that this position is not the C-terminus of the complete molecule; there is no description field for this key
NON_CONS	non consecutive residues; indicates that two residues in a sequence are not consecutive and that there are a number of unsequenced residues between them
UNSURE	uncertainties in the sequence; used to describe region(s) of a sequence for which the authors are unsure about the sequence assignment

**Appendix 3****Specimen sequence listing**

```

<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  cgcgctcctc  tcgcgcctct  ctctcgctct  cctctcgctc          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

```

<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



## 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
<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
<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 ; <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>- paires de bases (numéros des positions des paires de 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 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)
<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)

## 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	<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

Tableau 2 : liste des nucléotides modifiés

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-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
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

Tableau 3 : liste des acides aminés

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

Tableau 4 : liste des acides aminés modifiés ou peu connus

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
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

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 ("displacement loop") ; 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 ("diversity segment") 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	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 segment") 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 ("long terminal repeat") ; 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 répliquative, 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, scrRNA 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
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
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
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 auteurs signalent qu'il existe des variants de la séquence
VARSP LIC	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 glycérite 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 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)
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

### 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  cggccctct  cgcgctcctc  tcgcgctct  ctctcgctct  cctctcgctc           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
    
```

<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

## Mitteilung des Europäischen Patentamts vom 2. Oktober über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen

Für internationale Anmeldungen nach dem PCT gilt ab 1. Juli 1998 der "Standard über die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in internationalen Anmeldungen nach dem PCT"<sup>1</sup> (abgekürzt "PCT-Standard", siehe Regeln 5.2, 12.1 (d) und 13<sup>ter</sup> PCT sowie Abschnitt 208 und Anhang C der PCT-Verwaltungsvorschriften). Zum selben Zeitpunkt ist WIPO-Standard ST. 25 für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in Patentanmeldungen in Kraft getreten. Dieser neue WIPO-Standard, der die bisherigen WIPO-Standards ST. 23 und ST. 24 ersetzt, empfiehlt den Patentämtern, den PCT-Standard entsprechend auf alle Patentanmeldungen, bei denen es sich nicht um internationale Anmeldungen handelt, anzuwenden.

Gemäß Beschluß des Präsidenten des EPA vom 2. Oktober 1998<sup>2</sup> (nachfolgend "Beschluß des Präsidenten" genannt) ist der WIPO-Standard ST. 25 und damit der PCT-Standard für das europäische Patenterteilungsverfahren ab dem 1. Januar 1999 maßgebend.

Diese Mitteilung ergänzt den Beschluß des Präsidenten und erläutert das Verfahren nach dem EPÜ und dem PCT.

### I. Allgemeines

Seit dem 1. Januar 1993 ist in Verfahren vor dem EPA nach dem EPÜ und dem PCT die Darstellung von Nucleotid- und Aminosäuresequenzen in Form eines standardisierten Sequenzprotokolls sowie dessen Einreichung auf einem elektronischen Datenträger zwingend vorgeschrieben. Der neue WIPO Standard ST. 25 ersetzt die bisherigen Standards ST. 23 und ST. 24. Damit werden neben der Berücksichtigung des technologischen Fortschritts Vereinfachungen, Klarstellungen und Erleichterungen für die Anmelder

<sup>1</sup> Abgedruckt in der PCT-Gazette (Sonderausgabe) vom 25.6.1998.

<sup>2</sup> Abgedruckt in dieser Beilage auf Seite 1.

## Notice from the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings

The "Standard for the presentation of nucleotide and amino acid sequence listings in international applications under the PCT"<sup>1</sup> (or "PCT Standard", see Rules 5.2, 12.1(d) and 13<sup>ter</sup> PCT as well as Section 208 of and Annex C to the PCT Administrative Instructions) came into force for international applications under the PCT on 1 July 1998. WIPO Standard ST.25 for the presentation of nucleotide and amino acid sequences in patent applications also came into force on that date. This new WIPO Standard, which replaces WIPO Standards ST.23 and ST.24, recommends that patent offices also apply the PCT Standard mutatis mutandis to all patent applications in addition to international applications.

According to the decision of the President of the European Patent Office dated 2 October 1998 (hereinafter referred to as the "decision of the President"), WIPO Standard ST.25, and hence the PCT Standard, will enter into force for the European patent grant procedure on 1 January 1999.

This notice supplements the decision of the President and is intended to clarify the EPC and PCT procedures.

### I. General remarks

The presentation of nucleotide and amino acid sequences in the form of a standardised sequence listing as well as the filing thereof on an electronic data carrier has been mandatory in proceedings before the EPO under the EPC and the PCT since 1 January 1993. The new WIPO Standard ST.25 replaces Standards ST.23 and ST.24. It not only takes account of advances in technology, but also simplifies, clarifies and facilitates procedures for the applicant. In particular, it will no longer be necessary to translate a sequence

<sup>1</sup> See special edition of the PCT Gazette of 25 June 1998.

<sup>2</sup> See page 1 of this supplement.

## Communiqué de l'Office européen des brevets en date du 2 octobre 1998, relatif à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences

A compter du 1<sup>er</sup> juillet 1998, la norme applicable aux demandes internationales au titre du PCT est 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 PCT"<sup>1</sup> (en abrégé "norme du PCT", cf. règles 5.2, 12.1 d) et 13<sup>ter</sup> PCT ainsi que l'instruction administrative 208 et l'annexe C des Instructions administratives du PCT). A la même date est entrée en vigueur la norme ST. 25 de l'OMPI relative à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet. Cette nouvelle norme de l'OMPI, qui remplace les normes ST. 23 et ST. 24 de l'OMPI qui étaient applicables jusqu'ici, recommande aux offices de brevets d'appliquer par analogie la norme du PCT pour toutes les demandes de brevet autres que les demandes internationales.

Conformément à la décision du Président de l'OEB en date du 2 octobre 1998<sup>2</sup> (dénommée ci-après "décision du Président"), la norme ST. 25 de l'OMPI (et donc la norme du PCT) est applicable à compter du 1<sup>er</sup> janvier 1999 pour la procédure de délivrance de brevets européens.

Le présent communiqué complète la décision du Président et fournit des explications concernant la procédure selon la CBE et selon le PCT.

### I. Généralités

Dans les procédures au titre de la CBE ou du PCT se déroulant devant l'OEB, il est exigé impérativement depuis le 1<sup>er</sup> janvier 1993 la présentation d'un listage des séquences de nucléotides et d'acides aminés sous forme de listage des séquences normalisé, et le dépôt de ce listage sur un support électronique de données. La nouvelle norme ST. 25 de l'OMPI remplace les normes ST. 23 et ST. 24 qui s'appliquaient jusque là. Il est tenu compte ainsi du progrès technique, et de plus la situation se voit simplifiée, clarifiée et facilitée pour

<sup>1</sup> Reproduite dans la Gazette du PCT, numéro spécial du 25 juin 1998.

<sup>2</sup> Reproduite dans le présent supplément, page 1.

eingeführt. Insbesondere soll der Zwang zur Übersetzung eines Sequenzprotokolls bei Verwendung für verschiedene Anmeldungen in verschiedenen Sprachen dadurch entfallen, daß das Sequenzprotokoll unter Verzicht auf unnötige sprachliche Angaben (ersetzt durch numerische Codes) und unter Verwendung eines kontrollierten, von der Fachwelt allgemein anerkannten Vokabulars formuliert wird.

## II. Europäisches Patenterteilungsverfahren

### 1. Sequenzprotokoll

1.1 Die Darstellung von Nucleotid- und Aminosäuresequenzen im vorgeschriebenen Sequenzprotokoll muß dem im Anhang zum Beschluß des Präsidenten wiedergegebenen WIPO-Standard ST. 25 und damit dem PCT-Standard entsprechen (im folgenden "Standard" genannt).

1.2 Ein Sequenzprotokoll ist nur für solche Sequenzen einzureichen, die der Definition laut Nummer 2 (ii) des Standards entsprechen und sich auf die in der Anmeldung offenbarte Erfindung beziehen.

Soweit in der Beschreibung auf Sequenzen Bezug genommen wird, die zum Stand der Technik gehören und in allgemein zugänglichen Datenbanken auffindbar sind, ist es nicht erforderlich, solche Sequenzen in das Sequenzprotokoll aufzunehmen. Jedoch ist jede solche Sequenz hinreichend durch Angabe ihrer Datenbank-Zugangsnummer zu identifizieren.

1.3 Wird ein Sequenzprotokoll zusammen mit den Anmeldungsunterlagen, d. h. als Bestandteil der Beschreibung eingereicht, so sollen die Sequenzen nicht in der Beschreibung wiederholt werden (siehe Standard, Nummer 3).

1.4 Das Sequenzprotokoll kann entweder in der Sprache der Anmeldung oder (bevorzugt) in Englisch abgefaßt werden. Sind Sequenzprotokoll einerseits und Beschreibung und Ansprüche andererseits nicht in derselben Sprache abgefaßt und enthält das Sequenzprotokoll "freien Text" im Sinne des Standards (Nummer 33), so wird empfohlen, den freien Text am Ende der Beschreibung in der Sprache der Beschreibung gemäß Nummer 36 des Standards wiederzugeben (siehe Beispiele in der Anlage zu dieser Mitteilung). Damit kann vermieden werden, daß Merkmale, die nur im sogenannten "freien Text" des Sequenzprotokolls angegeben sind, in einzelnen Ver-

listing used for different applications in different languages, as sequence listings can be formulated without unnecessary text (replaced by numerical codes) and using a limited vocabulary of terms generally recognised in the field.

## II. European patent grant procedure

### 1. Sequence listing

1.1 The presentation of nucleotide and amino acid sequences in the prescribed sequence listing must comply with WIPO Standard ST.25, reproduced in the annex to the decision of the President, and thus with the PCT Standard (hereinafter referred to as the "Standard").

1.2 A sequence listing need only be filed for those sequences matching the definition given in the Standard, paragraph 2 (ii), and relating to the invention disclosed in the application.

Where the description refers to sequences which belong to the prior art and can be found in generally accessible databases, it will not be necessary to include such sequences in the sequence listing. However, each such sequence must be identified adequately by its database access number.

1.3 Where a sequence listing is filed together with the application documents, ie as part of the description, the sequences should not be repeated in the description (see Standard, paragraph 3).

1.4 The sequence listing may be written either in the language of filing or (preferably) in English. Where the sequence listing on the one hand and the description and claims on the other are not in the same language, and the sequence listing contains "free text" within the meaning of the Standard (paragraph 33), it is recommended that the free text be repeated at the end of the description in the language of the description in accordance with paragraph 36 of the Standard (see examples in the annex to this notice). This will ensure that features which are given only in the so-called "free text" of the sequence listing are not deemed in individual contracting states to the EPC not to

les demandeurs. Il n'y aura notamment plus lieu d'exiger la traduction d'un même listage des séquences figurant dans différentes demandes rédigées dans des langues différentes. Du fait que les indications linguistiques superflues disparaîtront du listage des séquences (elles seront remplacées par des codes numériques) et qu'il sera utilisé un vocabulaire contrôlé, universellement reconnu par les hommes du métier.

## II. Procédure de délivrance des brevets européens

### 1. Listage des séquences

1.1 La présentation du listage prescrit des séquences de nucléotides et d'acides aminés doit être conforme à la norme ST. 25 de l'OMPI reproduite dans l'annexe à la décision du Président, et donc à la norme du PCT (dénommée ci-après "la norme").

1.2 Il n'y a lieu de déposer de listage des séquences que pour les séquences correspondant à la définition figurant au paragraphe 2(ii) de la norme et se rapportant à l'invention exposée dans la demande.

Dans la mesure où il est fait référence dans la description à des séquences qui sont comprises dans l'état de la technique et que l'on peut trouver dans des banques de données accessibles au public, il n'est pas nécessaire d'intégrer ces séquences dans le listage des séquences. Toutefois il convient d'identifier suffisamment toutes ces séquences en indiquant leur numéro d'accès dans la banque de données.

1.3 Si un listage des séquences est déposé en même temps que les pièces de la demande, c'est-à-dire en tant que partie de la description, il n'y a pas lieu de reprendre les séquences dans la description (cf. norme, paragraphe 3).

1.4 Le listage des séquences peut être rédigé soit dans la langue de la demande, soit (de préférence) en anglais. Si le listage des séquences, d'une part, et la description et les revendications, d'autre part, ne sont pas rédigés dans la même langue, et que le listage des séquences ne comporte pas de "texte libre" au sens de la norme (paragraphe 33), il est recommandé de répéter le texte libre à la fin de la description, dans la langue de la description, conformément au paragraphe 36 de la norme, ce qui permettra d'éviter que dans certains Etats parties à la CBE des caractéristiques qui ne sont indiquées que dans le "texte libre" du listage des séquences ne soient pas



tragsstaaten des EPÜ aufgrund der Sprache nicht als Gegenstand der ursprünglichen Offenbarung angesehen werden.

## 2. Elektronischer Datenträger

2.1 Die Einreichung des Sequenzprotokolls auf einem elektronischen Datenträger zusätzlich zur schriftlichen Form ist vorgeschrieben. Die näheren Einzelheiten über die zugelassenen Datenträger, ihre Kennzeichnung, das Format der darauf gespeicherten Information und die erforderliche Hardware sind dem Standard (Nummern 39 ff.) zu entnehmen.

2.2 Das EPA empfiehlt die Verwendung der Software **Patentin**, die überarbeitet worden ist (Version 2.0) und den Anmeldern kostenlos zur Verfügung steht. Durch Verwendung dieser Software (in der jeweils gültigen Version) kann der Anmelder erreichen, daß die Formerfordernisse nach Regel 27a (1) EPÜ erfüllt sind. Die Software kann von der Website des EPA heruntergeladen werden und eignet sich für PCs, die mit dem Betriebssystem Windows (Version 3.1 oder höher) ausgestattet sind. Weitere Informationen sind erhältlich bei:

Europäisches Patentamt  
 Programm STRAND  
 Patentlaan 2  
 NL-2288 EE Rijswijk  
 Fax: (+31-70) 340 3992  
 E-mail: Patentin\_help.europatent.org

### 2.3 Beschädigter oder unlesbarer Datenträger

Ist der zusammen mit der Anmeldung eingereichte Datenträger beschädigt oder aus einem anderen Grund nicht lesbar, so wird der Anmelder aufgefordert, einen einwandfreien Datenträger innerhalb einer nicht verlängerbaren Frist von einem Monat ab Zustellung der entsprechenden Mitteilung nachzureichen.

### 3. Erklärung gemäß Regel 27a (2) EPÜ

Der vorgeschriebene Datenträger wird für die EDV-gestützte europäische oder internationale Recherche sowie, nach der Veröffentlichung der Anmeldung, für die laufende Aktualisierung der internationalen Patent-Sequenzdatenbank benötigt.

Die auf dem Datenträger gespeicherte Information muß mit dem schriftlichen Sequenzprotokoll über-

be the subject-matter of the original disclosure because of the language used.

## 2. Electronic data carriers

2.1 Sequence listings must be submitted on an electronic data carrier as well as in written form. Further details with regard to authorised data carriers, their identification, the format for the information recorded on them and the necessary hardware are set out in the Standard (paragraphs 39 ff).

2.2 The EPO recommends the use of **Patentin** software, which has been upgraded (version 2.0) and is available to applicants free of charge. By using the (applicable) version of this software, applicants can ensure that the physical requirements under Rule 27a(1) EPC are met. The software can be downloaded from the EPO's website and is suitable for use with Windows 3.1 or above. Further information can be obtained from:

European Patent Office  
 STRAND program  
 Patentlaan 2  
 NL-2288 EE Rijswijk  
 Fax: (+31-70) 340 3992  
 e-mail: Patentin\_help.europatent.org

### 2.3 Damaged or unreadable data carriers

If a data carrier submitted to the EPO is damaged, or unreadable for some other reason, the applicant will be invited to submit a flawless replacement within a non-extendable period of one month from notification of the communication.

### 3. Statement under Rule 27a(2) EPC

The prescribed data carrier is required for EDP-based European or international searches and, after publication of the application, for continuous updating of the international patent sequence database.

The information recorded on the data carrier must be identical to the written sequence listing. This must

considérées comme l'objet de la divulgation initiale, en raison de la différence existant entre la langue utilisée dans la description et celle utilisée dans le texte libre.

## 2. Support électronique de données

2.1 La production du listage des séquences sur un support électronique de données en plus du listage imprimé est obligatoire. Des précisions concernant les supports électroniques admis, leur identification, le format des informations y figurant et le matériel nécessaire sont données dans la norme (paragraphes 39 s.)

2.2 L'OEB recommande l'utilisation du logiciel **Patentin**, qui a été revu (version 2.0) et est mis gratuitement à la disposition des demandeurs. Le demandeur qui utilise ce logiciel (dans la version en vigueur) est en mesure de satisfaire aux conditions de forme prescrites à la règle 27bis (1) CBE. Le logiciel peut être téléchargé à partir du site web de l'OEB et convient pour les PC qui sont équipés du programme d'exploitation Windows (version 3.1 ou versions suivantes). Pour plus d'informations s'adresser à :

Office européen des brevets  
 Programme STRAND  
 Patentlaan 2  
 NL-2288 EE Rijswijk  
 Fax: (+31-70) 340 3992  
 E-mail : Patentin\_help.europatent.org

### 2.3 Support de données endommagé ou illisible

Si le support de données déposé en même temps que la demande est endommagé ou illisible pour quelque autre raison, le demandeur est invité à déposer un support de données en parfait état dans un délai d'un mois à compter de la signification de la notification correspondante ; ce délai ne peut être prolongé.

### 3. Déclaration prévue à la règle 27bis (2) CBE

Le support de données prescrit est nécessaire pour la recherche européenne ou internationale assistée par ordinateur ainsi que, après la publication de la demande, pour la mise à jour constante de la banque internationale de données relatives aux séquences contenues dans les brevets.

L'information figurant sur ce support doit être identique à celle que contient le listage imprimé des séquen-

einstimmen. Der Anmelder oder sein Vertreter hat dies sicherzustellen und eine entsprechende Erklärung gemäß Regel 27a (2) EPÜ zusammen mit dem Datenträger einzureichen (siehe auch Standard, Nummer 40). Eine solche Erklärung ist im Erteilungsantrag enthalten (EPA Form 1001, Feld 31)<sup>3</sup>. Sie kann als Mustererklärung auch bei Nachreichung des Datenträgers verwendet werden.

Stellt die Prüfungsabteilung fest, daß der eingereichte Datenträger das Identitätserfordernis nicht erfüllt und somit die Erklärung gemäß Regel 27a (2) EPÜ unrichtig ist, so kann dies zur Zurückweisung der Anmeldung nach Artikel 97 (1) EPÜ führen. Darüber hinaus kann ein anderer Stand der Technik einschlägig sein, als er der Prüfung zugrunde gelegt wurde. Dies kann zu einer anderen Beurteilung der Patentierbarkeitsvoraussetzungen durch die jeweils zuständigen Instanzen auch nach Patenterteilung führen.

#### 4. Nachreichung eines Sequenzprotokolls

4.1 Ist ein Sequenzprotokoll einzureichen, so ist es den Anmeldeunterlagen als Teil der Beschreibung beizufügen (siehe Standard, Nummer 3).

4.2 Gleichwohl nimmt das EPA auch nach dem Anmeldetag eingereichte Sequenzprotokolle entgegen, sofern der vorgeschriebene Datenträger und die Erklärungen gemäß Regel 27a (2) und (3) EPÜ vorliegen, weil dies für die Durchführung der Recherche und die Fortschreibung der internationalen Patent-Sequenzdatenbank sinnvoll ist.

4.3 Ein später eingereichtes Sequenzprotokoll ist jedoch nicht Bestandteil der Beschreibung und kann daher für die Bestimmung des ursprünglichen Offenbarungsgehalts der Anmeldung nicht herangezogen werden (Regel 27a (4) EPÜ).

4.4 Ist der Anmelder nicht imstande, das Sequenzprotokoll mit der Anmeldung einzureichen, sollte er es von sich aus baldmöglichst nachreichen und die weiteren Erfordernisse nach Regel 27a (2) und (3) EPÜ erfüllen (Einreichung des vorgeschriebenen Datenträgers und der vorgesehenen Erklärungen, siehe auch Punkt II.5). Andernfalls fordert ihn die Eingangsstelle auf, die festgestellten Mängel innerhalb einer **Frist von zwei Mona-**

be ensured by the applicant or his representative, who must file together with the data carrier a statement in accordance with Rule 27a(2) EPC (see also Standard, paragraph 40). The statement is contained in the Request for Grant (EPO Form 1001, section 31)<sup>3</sup>, and may also be used as a model statement when the data carrier is filed subsequently.

If the examining division finds that the data carrier submitted does not comply with the requirement that it be identical and the statement under Rule 27a (2) EPC is therefore incorrect, this may lead to refusal of the application under Article 97(1) EPC. Furthermore, a different prior art than that on which the examination was based may be relevant. This may lead to a different assessment by the relevant departments or courts of the conditions for patentability, even after a patent has been granted.

#### 4. Subsequent filing of a sequence listing

4.1 If a sequence listing has to be filed, it should be annexed to the application documents as part of the description (see Standard, paragraph 3).

4.2 Nevertheless, the EPO accepts sequence listings filed after the date of filing so long as the prescribed data carrier and the statements under Rule 27a(2) and (3) EPC have been received because they can be useful in the search and for updating the international patent sequence database.

4.3 A sequence listing filed later does not, however, form part of the description and cannot therefore be used in determining the original disclosed content of the application (Rule 27a(4) EPC).

4.4 If the applicant is unable to file the sequence listing with the description, he should file it without prompting as soon as possible thereafter and comply with the other requirements under Rule 27a(2) and (3) EPC (see also point II.5 for submission of the prescribed data carrier and statements). Otherwise, the Receiving Section will invite him to remedy the deficiencies within a **period of two months** (Article 91(2),

ces. Le demandeur ou son mandataire est tenu de prendre les mesures nécessaires à cet effet et de joindre au support de données une déclaration en ce sens, conformément à la règle 27bis (2) CBE (cf. également la norme, paragraphe 40). Une telle déclaration figure dans la requête en délivrance (OEB Form 1001, rubrique 31)<sup>3</sup>. Elle peut également être utilisée comme déclaration type en cas de production ultérieure du support de données.

Si la division d'examen constate que le support de données ne satisfait pas à l'exigence d'identité avec la liste écrite et que la déclaration selon la règle 27bis (2) CBE est donc inexacte, la demande peut être rejetée conformément à l'article 97(1) CBE. Il se peut en outre qu'un état de la technique autre que celui qui a été pris en compte pour l'examen soit pertinent et qu'il en résulte une appréciation divergente par les instances compétentes des conditions de brevetabilité même après la délivrance du brevet.

#### 4. Dépôt ultérieur d'un listage des séquences

4.1 Tout listage des séquences doit être joint aux pièces de la demande comme faisant partie de la description (cf. norme, paragraphe 3).

4.2 L'OEB accepte néanmoins les listages déposés après la date de dépôt, dans la mesure où le support de données prescrit et les déclarations selon la règle 27bis (2) et (3) CBE ont été produits, étant donné que ces informations sont utiles pour effectuer la recherche et actualiser la banque internationale de données relatives aux séquences contenues dans les brevets.

4.3 Un listage de séquences déposé ultérieurement ne fait toutefois pas partie de la description et ne peut donc servir à déterminer le contenu initial de la demande (règle 27bis (4) CBE).

4.4 Si le demandeur n'est pas en mesure de déposer le listage des séquences avec la demande, il a intérêt à en effectuer spontanément le dépôt ultérieurement dans les meilleurs délais, en veillant à remplir les autres conditions énoncées à la règle 27bis (2) et (3) CBE (production du support de données prescrit et des déclarations prévues; cf. également point II.5). Sinon, la section de dépôt l'invite à remédier dans un

<sup>3</sup> Abgedruckt im ABI. EPA 1997, 251.

<sup>3</sup> Published in OJ EPO 1997, 251.

<sup>3</sup> Reproduit au JO OEB 1997, 251.

ten zu beheben (Art. 91 (2), Regel 41 (1) EPÜ und Artikel 3 des Beschlusses des Präsidenten). Die Frist wird **einmal um höchstens zwei Monate** verlängert, sofern der Anmelder dies vor Ablauf der Frist beantragt. In begründeten Ausnahmefällen kann die Frist auch von Amts wegen verlängert werden, wenn nicht alle Mängel innerhalb der gesetzten Frist beseitigt worden sind. Wird innerhalb dieser Frist ein **beschädigter oder aus einem anderen Grund unlesbarer Datenträger** eingereicht, so wird dem Anmelder eine besondere, nicht verlängerbare Frist von einem Monat ab Zustellung der entsprechenden Mitteilung zur Nachreichung eines einwandfreien Datenträgers eingeräumt.

##### 5. Erklärung gemäß Regel 27a (3) EPÜ

Ein nachgereichtes Sequenzprotokoll darf nichts enthalten, das über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht, da die Recherche auf der Grundlage des nachgereichten Sequenzprotokolls durchgeführt wird. Das nachgereichte Sequenzprotokoll darf daher nur diejenige Sequenzinformation – in standardisierter Form – enthalten, die in den ursprünglichen Anmeldeunterlagen bereits vorhanden war. Es ist unbedingt darauf zu achten, daß das nachgereichte Sequenzprotokoll nicht mehr und nicht weniger Sequenzen als die in der ursprünglichen Beschreibung offenbarten Sequenzen enthält und die Numerierung dieser Sequenzen unverändert bleibt. Der Anmelder hat dies sicherzustellen und eine entsprechende Erklärung gemäß Regel 27a (3) EPÜ einzureichen. Es wird empfohlen, hierfür die in der Fußnote abgedruckte Mustererklärung<sup>4</sup> zu verwenden. Für den ebenfalls einzureichenden Datenträger gilt Punkt II.3.

##### 6. Rechtsverlust

Werden die Erfordernisse der Regel 27a EPÜ in Verbindung mit dem Beschluß des Präsidenten nicht rechtzeitig, gegebenenfalls nach Aufforderung der Eingangsstelle gemäß Regel 41 (1) EPÜ, erfüllt, so wird die Anmeldung zurückgewiesen (Art. 91 (1) b) und (3) EPÜ).

Der Anmelder kann die Weiterbehandlung der Anmeldung nach Artikel 121 EPÜ beantragen.

Rule 41(1) EPC and Article 3 of the decision of the President). This period may be extended **once only for a maximum of two months**, provided that the applicant requests such an extension before the period expires. In exceptional, substantiated cases, the period may be extended by the Receiving Section of its own motion where not all the deficiencies have been remedied within the prescribed period. If a **data carrier is filed which is damaged, or unreadable for some other reason**, the applicant will be granted a special, non-extendable period of one month from notification of the communication within which to file a flawless replacement.

##### 5. Statement under Rule 27a(3) EPC

A subsequently filed sequence listing may not include any matter which goes beyond the content of the application as filed, since the search is to be carried out on the basis of the subsequently filed sequence listing. The subsequently filed sequence listing may therefore contain only the sequence information – in a standardised form – that was already in the original application documents. The subsequently filed sequence listing should contain no more and no fewer sequences than disclosed in the original description, and their numbering should remain unchanged. This must be ensured by the applicant, who must file a statement to that effect in accordance with Rule 27a(3) EPC. It is recommended that the specimen statement in footnote<sup>4</sup> below be used for this purpose. Point II.3 is applicable for the data carrier which must also be submitted.

##### 6. Loss of rights

If the requirements laid down in Rule 27a EPC in conjunction with the decision of the President are not met in good time – if applicable, after an invitation from the Receiving Section under Rule 41(1) EPC – the application is refused (Art. 91(1)(b) and (3) EPC).

The applicant may request further processing of the application under Article 121 EPC.

**délai de deux mois** aux irrégularités constatées (art. 91(2), règle 41(1) CBE et art. 3 de la décision du Président). Le délai est prolongé **une seule fois de deux mois au maximum**, si le demandeur présente une requête en ce sens avant l'expiration du délai. Dans des cas exceptionnels, si la situation le justifie, le délai peut également être prolongé d'office, s'il n'a pas été remédié à toutes les irrégularités dans le délai fixé. S'il est déposé dans ce délai un **support de données endommagé ou rendu illisible pour quelque autre raison**, il est imparti au demandeur un délai spécial non prorogeable d'un mois à compter de la signification de la notification correspondante invitant le demandeur à déposer un support de données en parfait état.

##### 5. Déclaration prévue à la règle 27bis (3) CBE

Un listage des séquences déposé ultérieurement ne doit contenir aucun élément s'étendant au-delà du contenu de la demande telle qu'elle a été déposée, puisque la recherche est exécutée sur la base de ce listage. Le listage des séquences déposé ultérieurement ne doit contenir – sous forme normalisée – que les informations concernant les séquences qui figuraient déjà dans les pièces initiales de la demande. Il faut absolument veiller à ce que le listage des séquences déposé ultérieurement ne contienne ni plus ni moins de séquences que celles qui sont divulguées dans la description initiale et à ce que la numérotation de ces séquences demeure inchangée. Le demandeur est tenu de faire le nécessaire à cet effet et de produire une déclaration en ce sens conformément à la règle 27bis (3) CBE. Il est recommandé d'utiliser à cette fin la déclaration type reproduite dans la note en bas de page<sup>4</sup>. Le point II.3 est applicable au support de données que le demandeur doit en outre produire.

##### 6. Perte de droits

Si le demandeur ne satisfait pas en temps utile aux conditions énoncées à la règle 27bis CBE en liaison avec la décision du Président de l'OEB, le cas échéant après y avoir été invité par la section de dépôt conformément à la règle 41 (1) CBE, la demande est rejetée (art. 91(1)(b) et (3) CBE).

Le demandeur peut demander la poursuite de la procédure conformément à l'article 121 CBE.

<sup>4</sup> "Es wird hiermit erklärt, daß das Sequenzprotokoll nicht über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht."

<sup>4</sup> "It is hereby declared that the sequence listing does not go beyond the content of the application as filed."

<sup>4</sup> "Il est déclaré par la présente que le listage des séquences ne s'étend pas au-delà du contenu de la demande telle qu'elle a été déposée."

### III. Verfahren vor dem EPA als Internationaler Recherchenbehörde oder als mit der internationalen vorläufigen Prüfung beauftragter Behörde

1. Sequenzprotokolle in maschinenlesbarer Form werden für die Durchführung der internationalen Recherche ebenso benötigt wie für die europäische Recherche. Sie müssen dem PCT-Standard entsprechen (Regel 5.2 PCT sowie Abschnitt 208 und Anhang C der PCT-Verwaltungsvorschriften).

Liegt ein diesem Standard entsprechendes Sequenzprotokoll zu Beginn der internationalen Recherche nicht oder nicht auf elektronischem Datenträger vor, so wird der Anmelder vom EPA aufgefordert, das Sequenzprotokoll und/oder den vorgeschriebenen Datenträger innerhalb einer **nicht verlängerbaren Frist von maximal zwei Monaten** nachzureichen (Regel 13<sup>ter</sup>.1 a) PCT). Diese wird wegen der engen Fristen der Regel 42 PCT für die Erstellung des Recherchenberichts regelmäßig nur **einen Monat** betragen.

2. Bei Nachreichung des Sequenzprotokolls und/oder des Datenträgers hat der Anmelder zu erklären, daß das nachgereichte Sequenzprotokoll nicht über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht und/oder die auf dem Datenträger gespeicherte Information mit dem schriftlichen Sequenzprotokoll übereinstimmt (Art. 4 des Beschlusses des Präsidenten und Standard, Nummern 37 und 40).

3. Werden diese Erfordernisse nicht oder nicht rechtzeitig erfüllt, so führt das EPA keine internationale Recherche durch, soweit eine sinnvolle Recherche nicht möglich ist (Regel 13<sup>ter</sup>.1 c) PCT). In diesem Fall wird der internationale Recherchenbericht ganz oder teilweise durch die Erklärung gemäß Artikel 17 (2) PCT ersetzt.

4. Ist der internationale Recherchenbericht aus den vorgenannten Gründen nicht oder nur teilweise erstellt worden, so führt das EPA als mit der internationalen vorläufigen Prüfung beauftragte Behörde keine internationale vorläufige Prüfung durch, soweit eine sinnvolle Prüfung nicht möglich ist (Regel 66.1 e) PCT). Dies wird dem Anmelder mitgeteilt und im internationalen vorläufigen Prüfungsbericht vermerkt (Regel 66.2 a) vii)

### III. Proceedings before the EPO as International Searching Authority or International Preliminary Examining Authority

1. Sequence listings in machine readable form are just as necessary in carrying out the international search as they are for the European search. They must comply with the PCT Standard (Rule 5.2 PCT as well as Section 208 of and Annex C to the PCT Administrative Instructions).

If at the start of the international search a sequence listing complying with this Standard has not been received or has not been received on an electronic data carrier, the EPO invites the applicant to file the sequence listing and/or the prescribed data carrier within a **non-extendable period of not more than two months** (Rule 13<sup>ter</sup>.1(a) PCT). Because of the short time limits in Rule 42 PCT for the establishment of the search report, this period will normally be only **one month**.

2. If the sequence listing and/or data carrier is filed subsequently, the applicant must state that the sequence listing so filed does not include matter which goes beyond the content of the application as filed and/or that the information recorded on the data carrier is identical to the written sequence listing (Article 4 of the decision of the President, and Standard, paragraphs 37 and 40).

3. If these requirements are not met or not met in due time, the EPO does not perform the international search where a meaningful search cannot be carried out (Rule 13<sup>ter</sup>.1(c) PCT). In this case the international search report is replaced in full or in part by the statement under Article 17(2) PCT.

4. If the international search report has not been drawn up in full or in part for the aforementioned reasons, the EPO as International Preliminary Examining Authority does not perform the international preliminary examination where a meaningful examination cannot be carried out (Rule 66.1(e) PCT). This is notified to the applicant and stated in the international preliminary examination report (Rules 66.2(a)(vii) and 70.12(iv))

### III. Procédure devant l'OEB agissant en qualité d'administration chargée de la recherche internationale ou d'administration chargée de l'examen préliminaire international

1. Les listages des séquences sous forme déchiffirable par machine sont nécessaires tant pour la recherche internationale que pour la recherche européenne. Ils doivent être établis conformément à la norme du PCT (règle 5.2 PCT ainsi que instruction administrative 208 et annexe C des Instructions administratives du PCT).

Si un listage des séquences conforme à cette norme n'est pas parvenu à l'OEB ou bien s'il n'est pas disponible sur support électronique de données lorsque la recherche internationale est entreprise, le demandeur est invité par l'OEB à fournir **dans un délai non prorogable de deux mois au maximum** le listage et/ou le support de données prescrit (règle 13<sup>ter</sup>.1 a) PCT). Ce délai ne sera normalement que d'**un mois** en raison des délais très stricts fixés par la règle 42 PCT pour l'établissement du rapport de recherche internationale.

2. Lorsqu'il fournit ultérieurement le listage des séquences et/ou le support de données, le demandeur est tenu de déclarer que ce listage ne s'étend pas au-delà du contenu de la demande telle qu'elle a été déposée et/ou que l'information figurant sur ce support est identique à celle qui contient le listage imprimé (art. 4 de la décision du Président et norme, paragraphes 37 et 40).

3. Si ces conditions ne sont pas remplies ou ne le sont pas dans les délais, l'OEB ne procède pas à la recherche internationale dans la mesure où une recherche significative ne peut pas être effectuée (règle 13<sup>ter</sup>.1 c) PCT). Dans ce cas, le rapport de recherche internationale est remplacé en tout ou en partie par la déclaration prévue à l'article 17(2) PCT.

4. Si le rapport de recherche internationale n'a pas été établi, ou n'a été établi qu'en partie, pour les raisons qui viennent d'être énoncées, l'OEB agissant en qualité d'administration chargée de l'examen préliminaire international ne procède pas à l'examen préliminaire international dans la mesure où un examen significatif ne peut pas être effectué (règle 66.1 e) PCT). L'OEB en informe le demandeur et l'indique dans le rapport

und 70.12 iv) PCT). Eine Aufforderung zur Einreichung eines Sequenzprotokolls gemäß Regel 13<sup>ter</sup>.1 (e) PCT ergeht nicht.

Es wird nachdrücklich davon abgeraten, ein Sequenzprotokoll erst in diesem späten Stadium einzureichen. Zumindest bei umfangreichen Sequenzen wäre das EPA nicht in der Lage zu überprüfen, ob das Sequenzprotokoll über den Offenbarungsgehalt der internationalen Anmeldung im Anmeldezeitpunkt hinausgeht. Zu einer derartigen Überprüfung ist es nach Artikel 34 (2) b) und Regel 66.2 (a) iv) PCT zwar verpflichtet. Anders als bei der internationalen Recherche kann eine Erklärung des Anmelders, daß das nachgereichte Sequenzprotokoll nicht über den Offenbarungsgehalt der internationalen Anmeldung hinausgeht, das EPA nicht von einer eigenen Überprüfung dieses Tatbestandes (Artikel 34 (2) b), Regel 66.2 a) PCT) entbinden. Es gibt aber derzeit keine geeigneten Mittel, die einen zufriedenstellenden Vergleich der ursprünglichen Anmeldungsunterlagen mit dem nachgereichten Sequenzprotokoll zu vertretbarem Aufwand ermöglichen könnten. Daher könnte der internationale vorläufige Prüfungsbericht nicht auf der Grundlage des nachgereichten Sequenzprotokolls, sondern nur auf der Grundlage der am Anmeldetag in der Beschreibung vorhandenen Sequenzen erstellt werden.

#### IV. Eintritt der Euro-PCT-Anmeldungen in die regionale Phase vor dem EPA

1. Liegt ein nach Regel 5.2 PCT vorgeschriebenes Sequenzprotokoll bei Ablauf der internationalen Phase dem EPA als Bestimmungsamt oder ausgewähltem Amt nicht vor, entspricht es nicht dem Standard oder ist der vorgeschriebene Datenträger (siehe Standard, Nummern 39 ff.) nicht eingereicht worden, so wird der Anmelder aufgefordert, die festgestellten Mängel innerhalb einer **Frist von zwei Monaten** zu beheben (Regel 13<sup>ter</sup>.2 PCT, Regel 104b (3a) EPÜ und Artikel 5 des Beschlusses des Präsidenten). Gegebenenfalls wird der Anmelder aufgefordert, gleichzeitig zu erklären, daß das nachgereichte Sequenzprotokoll nicht über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht und/oder die auf dem

PCT). An invitation under Rule 13<sup>ter</sup>.1(e) PCT to file a sequence listing is not issued.

Applicants are strongly advised not to file sequence listings at this late stage. In the case of long sequence listings at least, the European Patent Office would not be in a position to examine whether the sequence listing goes beyond the content of the international application as disclosed at the time of filing, although under Article 34(2)(b) and Rule 66.2(a)(iv) PCT it is obliged to carry out such an examination. Unlike the situation pertaining to the international search, a statement from the applicant to the effect that the subsequently filed sequence listing does not go beyond the content of the international application as filed cannot release the EPO from its obligation to carry out its own examination (Article 34(2)(b), Rule 66.2(a) PCT). However, there is at present no appropriate means of carrying out a satisfactory comparison of the documents as originally filed with the subsequently filed sequence listing. The international preliminary examination report therefore could not be established on the basis of the subsequently filed sequence listing but only on the basis of the sequences available in the description on the date of filing.

#### IV. Entry of Euro-PCT applications into the regional phase before the EPO

1. If on expiry of the international phase the sequence listing prescribed in Rule 5.2 PCT is not available to the EPO as designated Office or elected Office or does not comply with the Standard, or if the prescribed data carrier (see Standard, paragraphs 39 ff) has not been filed, the applicant is invited to remedy such deficiencies within a **period of two months** (Rule 13<sup>ter</sup>.2 PCT, Rule 104b(3a) EPC and Article 5 of the decision of the President). Where appropriate, the applicant is invited at the same time to state that the subsequently filed sequence listing does not go beyond the content of the application as filed and/or that the information recorded on the data carrier is identical to the written sequence listing (Standard, para-

d'examen préliminaire international (règles 66.2 a) vii) et 70.12 iv) PCT). Il n'est pas envoyé d'invitation à déposer un listage des séquences conformément à la règle 13<sup>ter</sup>.1 e) PCT.

Il est tout à fait déconseillé de ne déposer un listage des séquences qu'à ce stade tardif de la procédure. Tout au moins dans le cas de séquences longues, l'OEB ne serait pas en mesure de vérifier si le listage des séquences ne s'étend pas au-delà de l'exposé de l'invention figurant dans la demande internationale à la date du dépôt. Il est certes tenu de procéder à une telle vérification en vertu de l'article 34 (2) b) et de la règle 66.2 a) iv) PCT, mais contrairement à ce qui se passe dans le cas de la recherche internationale, il ne suffit pas que le demandeur déclare que le listage des séquences déposé ultérieurement ne s'étend pas au-delà de l'exposé de l'invention figurant dans la demande internationale à la date du dépôt pour que l'OEB puisse se dispenser de procéder lui-même à cette vérification (article 34 (2) b), règle 66.2 a) PCT). Toutefois il n'existe pas actuellement de moyen approprié permettant d'établir une comparaison satisfaisante, ne nécessitant pas une somme de travail excessive, entre les pièces initiales de la demande et le listage des séquences déposé ultérieurement. C'est pourquoi le rapport d'examen préliminaire international ne saurait être établi sur la base du listage des séquences déposé ultérieurement, mais uniquement sur la base des séquences figurant dans la description à la date du dépôt.

#### IV. Entrée des demandes euro-PCT dans la phase régionale devant l'OEB

1. Si, au terme de la phase internationale, le listage des séquences visé à la règle 5.2 PCT n'est pas parvenu à l'OEB agissant en qualité d'Office désigné ou d'Office élu, n'a pas été établi conformément à la norme ou n'a pas été produit sur le support de données prescrit (cf. norme, paragraphes 39 s.), le demandeur est invité à remédier dans un **délai de deux mois** aux irrégularités constatées (règle 13<sup>ter</sup>.2 PCT, règle 104<sup>ter</sup> (3bis) CBE et article 5 de la décision du Président). Le cas échéant, le demandeur est invité à déclarer en même temps que le listage des séquences déposé ultérieurement ne s'étend pas au-delà du contenu de la demande telle qu'elle a été déposée et/ou que l'information figurant sur le support de données est identique à

Datenträger gespeicherte Information mit dem schriftlichen Sequenzprotokoll übereinstimmt (Standard, Nummern 37 und 40). Artikel 3 (2) und (3) des Beschlusses des Präsidenten gelten entsprechend (vgl. oben Punkt II 4.4).

2. Werden bestehende Mängel nach einer solchen Aufforderung nicht oder nicht rechtzeitig beseitigt, so wird die Anmeldung zurückgewiesen (Art. 91 (3) EPÜ).

3. Der Anmelder kann die Weiterbehandlung der Anmeldung nach Artikel 121 EPÜ beantragen.

4. Übersetzung gemäß Artikel 158 (2) EPÜ

Einer Übersetzung der Textstellen eines Sequenzprotokolls, das Bestandteil der Beschreibung ist, bedarf es grundsätzlich nicht, sofern das Sequenzprotokoll der Regel 12.1 (d) PCT und die Beschreibung der Regel 5.2 (b) PCT entspricht (Regel 49.5 (a-bis) PCT). Ist die Voraussetzung nach Regel 5.2 (b) PCT nicht erfüllt, so verlangt das EPA keine solche Übersetzung, wenn das Sequenzprotokoll in englisch abgefaßt ist und die internationale Anmeldung in einer anderen EPA-Amtssprache (deutsch oder französisch) veröffentlicht worden ist.

#### V. Veröffentlichung von Sequenzprotokollen durch das EPA – Akteneinsicht

1. Ist das Sequenzprotokoll **zusammen** mit den Unterlagen der europäischen Patentanmeldung (oder internationalen Anmeldung) als Teil der Beschreibung eingereicht worden, so wird es zusammen mit der Anmeldung durch das EPA (oder durch das Internationale Büro) veröffentlicht.

2. Wird ein Sequenzprotokoll zu einer europäischen oder zu einer in der regionalen Phase vor dem EPA anhängigen internationalen Anmeldung **nachgereicht**, so hängt seine Veröffentlichung von dem Zeitpunkt der Einreichung ab (Artikel 6 des Beschlusses des Präsidenten):

Wird es rechtzeitig vor Abschluß der technischen Vorbereitungen für die Veröffentlichung der europäischen Patentanmeldung<sup>5</sup> eingereicht, so wird es als Anlage zu den Anmeldeunterlagen (A-Schrift) oder im Falle einer Euro-PCT-Anmeldung, die nicht in einer Amtssprache des EPA

graphs 37 and 40). Article 3(2) and (3) of the decision of the President applies mutatis mutandis (see point II 4.4 above).

2. If existing deficiencies are not remedied or not remedied in due time after such an invitation, the application is refused (Art. 91(3) EPC).

3. The applicant may request further processing of the application under Article 121 EPC.

4. Translation in accordance with Article 158(2) EPC

There is no requirement to translate any text matter in a sequence listing which forms part of the description, provided that the sequence listing complies with Rule 12.1(d) PCT and the description with Rule 5.2(b) PCT (Rule 49.5(a-bis) PCT). If the requirement under Rule 5.2(b) PCT is not met, the EPO does not require such a translation where the sequence listing is in English and the international application has been published in one of the other official languages of the EPO (French or German).

#### V. Publication of sequence listings by the EPO – inspection of files

1. If the sequence listing has been filed **together** with the documents making up the European (or international) application as part of the description, it is published together with the application by the EPO (or the International Bureau).

2. If a sequence listing pertaining to a European patent application or to an international application in the regional phase before the EPO is **subsequently filed**, its publication depends on the date of filing (Article 6 of the decision of the President):

If it is filed in due time before completion of the technical preparations for publication of the European patent application<sup>5</sup>, it is published as an annex to the application documents (A document) or, in the case of a Euro-PCT application which has not been published internationally in one

celle que contient le listage imprimé (la norme, paragraphes 37 et 40). L'article 3(2) et (3) de la décision du Président est applicable par analogie (cf. supra, point II 4.4).

2. Si, après une telle invitation, le demandeur ne remédie pas aux irrégularités constatées ou n'y remédie pas en temps utile, la demande est rejetée (art. 91(3) CBE).

3. Le demandeur peut demander la poursuite de la procédure conformément à l'article 121 CBE.

4. Traduction conformément à l'article 158(2) CBE

Il n'est pas nécessaire en tout état de cause de déposer une traduction des éléments de texte d'un listage des séquences qui fait partie de la description, si ce listage des séquences est conforme à la règle 12.1 d) PCT et si la description est conforme à la règle 5.2 b) PCT (règle 49.5 (a-bis) PCT). Si la condition requise à la règle 5.2 b) PCT n'est pas remplie, l'OEB n'exige pas une telle traduction si le listage des séquences est rédigé en anglais et si la demande internationale a été publiée dans une autre langue officielle de l'OEB (allemand ou français).

#### V. Publication par l'OEB des listages des séquences – inspection publique

1. Si le listage des séquences a été déposé **en même temps** que les pièces de la demande européenne ou internationale comme partie de la description, il est publié avec la demande par l'OEB (ou par le Bureau international).

2. Si le listage des séquences d'une demande européenne ou d'une demande internationale se trouvant en instance devant l'OEB dans la phase régionale a été **déposé ultérieurement**, sa publication dépend de la date de son dépôt (article 6 de la décision du Président).

S'il est déposé en temps utile avant la fin des préparatifs techniques entrepris en vue de la publication de la demande de brevet européen<sup>5</sup>, il est publié en annexe aux pièces de la demande (document A) ou, s'il s'agit d'une demande euro-PCT qui n'a pas fait l'objet d'une publication interna-

<sup>5</sup> 7 Wochen vor Ablauf von 18 Monaten nach dem Anmelde- oder Prioritätstag; Beschluß des Präsidenten des EPA vom 14. Dezember 1992, ABI. EPA 1993, 55.

<sup>5</sup> Seven weeks before expiry of the eighteenth month following the date of filing or priority; Decision of the President of the EPO dated 14 December 1992, OJ EPO 1993, 55.

<sup>5</sup> Sept semaines avant l'expiration d'un délai de dix-huit mois à compter de la date de dépôt ou de priorité ; décision du Président de l'OEB en date du 14 décembre 1992, JO OEB 1993, 55.

international veröffentlicht worden ist, als Anlage zu der nach Artikel 158 (3) EPÜ veröffentlichten Übersetzung mitveröffentlicht.

In allen anderen Fällen besteht nur die Möglichkeit, bei Patenterteilung das Sequenzprotokoll zusammen mit der Patentschrift zu veröffentlichen. Die Veröffentlichung des nachgereichten Sequenzprotokolls zusammen mit dem gesonderten Recherchenbericht (sogenannte A3-Schrift) ist nicht vorgesehen.

3. Ein Sequenzprotokoll, das nicht zusammen mit den Anmeldeunterlagen oder mit der Patentschrift veröffentlicht wird, ist gegebenenfalls nur durch Akteneinsicht oder über die Patent-Sequenzdatenbanken der Öffentlichkeit zugänglich. Auf Antrag wird Akteneinsicht auch durch Herstellung und Zusendung einer Kopie des vom Anmelder eingereichten elektronischen Datenträgers gewährt (siehe II.2 und IV.1 und Regel 94 (3) EPÜ).

of the official languages of the EPO, as an annex to the translation published under Article 158(3) EPC.

In all other cases it is only possible to publish the sequence listing together with the patent specification when the patent is granted. There is no provision for publishing the subsequently filed sequence listing together with the separate search report (A3 document).

3. A sequence listing which is not published together with the application documents or with the patent specification is, where applicable, accessible to the public only by way of a file inspection or by way of the patent sequence databases. On request, inspection of files may also be granted by sending the requester a copy of the electronic data carrier filed by the applicant (see II.2 and IV.1 and Rule 94(3) EPC).

tionale dans une langue officielle de l'OEB, en annexe à la traduction publiée conformément à l'article 158(3) CBE.

Dans tous les autres cas, la seule possibilité consiste, si la demande donne lieu à la délivrance d'un brevet, à publier le listage des séquences avec le fascicule de brevet. Il n'est pas prévu qu'un listage des séquences déposé ultérieurement soit publié avec le rapport de recherche séparé (document A3).

3. Un listage des séquences qui n'est publié ni avec les pièces de la demande ni avec le fascicule de brevet n'est le cas échéant accessible au public que dans le cadre de l'inspection publique ou par l'intermédiaire de la banque de données relatives aux séquences contenues dans les brevets. Il est également procédé, sur requête, à l'inspection publique, par établissement et envoi d'une copie du support de données électronique produit par le demandeur (cf. II.2 et IV.1 et règle 94(3) CBE).

**Anlage zu der Mitteilung des Europäischen Patentamts vom 2. Oktober 1998 über die Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen und die Einreichung von Sequenzprotokollen\***

**Appendix to the Notice from the European Patent Office dated 2 October 1998 concerning the presentation of nucleotide and amino acid sequences in patent applications and the filing of sequence listings\***

**Annexe au Communiqué de l'Office européen des brevets en date du 2 octobre 1998, relatif à la présentation du listage des séquences de nucléotides et d'acides aminés dans les demandes de brevet et au dépôt de listages des séquences\***

**Freier Text des Sequenzprotokolls**

**Sequence Listing Free Text**

**Texte libre du listage des séquences**

**Beispiel 1:**

**Example 1:**

**Exemple 1 :**

Dies ist ein Beispiel des Abschnitts "Freier Text des Sequenzprotokolls" (vgl. Nummer 36 des WIPO-Standards ST. 25) in der Beschreibung einer in **Französisch** eingereichten europäischen oder internationalen Patentanmeldung, die ein Sequenzprotokoll mit Freitextelementen in Englisch zur Sequenz Nr. 3 enthält:

This is an example of a "Sequence Listing Free Text" section (see paragraph 36 of WIPO Standard ST. 25) in the description of a European patent application or international application filed in **French** and containing a sequence listing with English free text data elements for Sequence No. 3:

Ceci est un exemple de la rubrique particulière de la description intitulée "texte libre du listage des séquences" (voir le paragraphe 36 de la norme OMPI ST.25) pour une demande de brevet européen ou une demande internationale déposée en **français** et contenant un listage des séquences comportant des éléments de texte libre en anglais pour la séquence n° 3 :

```
<210> 3
<223> /note="PCV"
        /hôte spécifique="Sus scrofa"
        /type de tissu="rein"
        /lignée cellulaire="PK-15"
<223> /note="TAGTATTAC homologue au circovirus de plante et au motif nonanucléotide de géminivirus"
<223> /Codon de départ =1
        /note="similaire au circovirus BBTV de plante"
        /produit="site de réplication de la protéine associée au PCV"
<223> /note="fin de réplication de la protéine associée au PCV a lieu 6 nucléotides en aval du signal de polyadénylation"
<223> /note="TAGTATTAC homologue au circovirus de plante et au motif nonanucléotide de géminivirus"
```

**Beispiel 2:**

**Example 2**

**Exemple 2 :**

Dies ist ein Beispiel des Abschnitts "Freier Text des Sequenzprotokolls" (vgl. Nummer 36 des WIPO-Standards ST. 25) in der Beschreibung einer in **Deutsch** eingereichten europäischen oder internationalen Patentanmeldung, die ein Sequenzprotokoll mit Freitextelementen in Englisch zur Sequenz Nr. 3 enthält:

This is an example of a "Sequence Listing Free Text" section (see paragraph 36 of WIPO Standard ST. 25) in the description of a European patent application or international application filed in **German** and containing a sequence listing with English free text data elements for Sequence No. 3:

Ceci est un exemple de la rubrique particulière de la description intitulée "texte libre du listage des séquences" (voir le paragraphe 36 de la norme OMPI ST.25) pour une demande de brevet européen ou une demande internationale déposée en **allemand** et contenant un listage des séquences comportant des éléments de texte libre en anglais pour la séquence n° 3 :

```
<210> 3
<223> /Anmerkung="PCV"
        /Spezifischer Wirt="Sus scrofa"
        /Geweebetyp="Niere"
        /Zelllinie="PK-15"
<223> /Anmerkung="TAGTATTAC-Sequenz, homolog zu pflanzlichem Circovirus- und Geminivirus-spezifischem Nona-Nucleotidmotiv"
<223> /Startcodon = 1
        /Anmerkung="ähnlich dem pflanzlichen Circovirus BBTV"
        /Produkt="Replikationsstelle des assoziierten Proteins PCV"
<223> /Anmerkung="Ende des mutmaßlichen replikations-assozierten Proteins PCV tritt 6 Nucleotide stromabwärts des potentiellen Polyadenylierungssignals auf"
<223> /Anmerkung="TAGTATTAC-Sequenz, homolog zu pflanzlichem Circovirus- und Geminivirus-spezifischem Nona-Nucleotidmotiv"
```

\* Siehe II 1.4 der Mitteilung in dieser Beilage, S. 60.

\* See II 1.4 of the Notice in this supplement, p. 60.

\* Voir le point II 1.4 du communiqué reproduit dans le présent supplément, p. 60.





