

# AMTSBLATT DES EUROPÄISCHEN PATENTAMTS

18. Dezember 1989

## BEILAGE ZUM AMTSBLATT 12/1989

### Mitteilung des Europäischen Patentamts vom 15. November 1989 über die Darstellung von Nucleotid- und Aminosäuresequenzen In Patentanmeldungen

Das EPA hat Überlegungen angestellt, wie die Veröffentlichung und die Recherche europäischer Patentanmeldungen, die Nuclein- und Aminosäuresequenzen enthalten, vereinfacht werden könnten. In diesem Zusammenhang wird den Anmeldern ab **1. April 1990** dringend empfohlen, sich bei der Einreichung von Patentanmeldungen mit solchen Sequenzen an bestimmte Standardsymbole und Darstellungsverfahren zu halten. Bis zur entsprechenden Änderung der Ausführungsordnung zum EPÜ dient diese Empfehlung als Übergangsmaßnahme.

#### 1. Grundkonzept

Auf dem Gebiet der Biotechnologie hat mit der enormen Flut wissenschaftlicher Veröffentlichungen in den vergangenen Jahren auch der Ausstoß an Patentanmeldungen beträchtlich zugenommen. Der Umfang dieser Daten übersteigt das kognitive Erinnerungs-, Vergleichs- und Erkenntnisvermögen des Menschen.

Überdies gestalten sich die Recherche und Analyse der Patentanmeldungen, die Nucleinsäure-, d. h. DNA/RNA-Sequenzen und Aminosäure-, d. h. Proteinsequenzen betreffen und genetische Informationen in codierter Form enthalten, aufgrund ihrer hohen Komplexität selbst unter optimalen Bedingungen ausgesprochen schwierig. Festzustellen, ob zwei Sequenzen verwandt sind, und sie miteinander zu vergleichen, ist keineswegs einfach.

Derzeit gibt es keine Normen für die Wiedergabe von Sequenzdaten in Patentanmeldungen. Sequenzen werden in vielfältiger Weise mit nicht standardisierten Symbolen (einbuchstabiger oder

# OFFICIAL JOURNAL OF THE EUROPEAN PATENT OFFICE

18 December 1989

## SUPPLEMENT TO OFFICIAL JOURNAL 12/1989

### Notice from the European Patent Office dated 15 November 1989 concerning the representation of nucleotide and amino acid sequences In patent applications

The EPO has been looking at ways in which the publication and searching of European patent applications containing nucleic acid and amino acid sequences might be simplified. From **1 April 1990** therefore it will be strongly recommended that patent applications containing such sequences be filed using a set of standard symbols and representation methods. This recommendation will be temporary pending an amendment to the EPC Implementing Regulations.

#### 1. Basic concept

In the field of biotechnology the influx of patent applications has grown considerably over the past few years, accompanied by a staggering increase in scientific publications. The volume of these data exceeds the human cognitive capacity for remembering, comparing and drawing inferences.

Moreover, the sheer complexity of patent applications dealing with nucleic acid (DNA/RNA) and amino acid (protein) sequences comprising encoded forms of genetic information makes their search and analysis very difficult even under optimum conditions. Establishing whether two sequences are related and comparing them is not an easy matter.

Currently there are no standards for the representation of sequence data in patent applications. Applicants have recourse to a variety of sequence descriptions and representations: non-standard-

# JOURNAL OFFICIEL DE L'OFFICE EUROPÉEN DES BREVETS

18 décembre 1989

## SUPPLEMENT AU JOURNAL OFFICIEL n° 12/1989

### Communiqué de l'Office européen des brevets, en date du 15 novembre 1989, relatif à la représentation de séquences nucléotidiques et d'acides aminés dans des demandes de brevet

L'OEB a recherché des moyens de simplifier la publication des demandes de brevet européen traitant de séquences d'acides nucléiques et d'acides aminés et la recherche s'y rapportant. A compter du **1er avril 1990**, il sera vivement recommandé de déposer les demandes de brevet concernant de telles séquences en utilisant un ensemble de symboles normalisés et de méthodes de représentation. Cette recommandation vaudra jusqu'à ce que le règlement d'exécution de la CBE soit modifié en conséquence.

#### 1. Conception de base

Parallèlement à l'accroissement spectaculaire des publications scientifiques, le nombre de demandes de brevet déposées en biotechnologie a considérablement augmenté au cours de ces quelques dernières années. Le volume de ces données dépasse les facultés humaines de mémorisation, de comparaison et d'analyse.

En outre, la grande complexité des demandes de brevet traitant de séquences d'acides nucléiques (ADN/ARN) et d'acides aminés (protéines) comprenant des informations génétiques sous forme codée rend la recherche et l'analyse très difficiles, même dans des conditions optimales. Il n'est pas simple de déterminer si deux séquences sont connexes et de les comparer.

A l'heure actuelle, il n'existe pas de normes de représentation des données relatives aux séquences dans les demandes de brevet. Les demandeurs décrivent et représentent les séquences

dreibuchstabiger Code für Aminosäuren), anhand einzel- oder doppelsträngiger DNA, unter Heranziehung der entsprechenden RNA, mit willkürlicher Nummerierung, ungenauen und mehrdeutigen Angaben manipulierter Sequenzteile usw. beschrieben und dargestellt. Die Tatsache, daß Symbole nicht konsequent verwendet werden und ein Standardformat für Sequenzen fehlt, erschwert die ohnehin mühsame Recherche und Analyse zusätzlich; im Interesse der Prüfer wie auch der Öffentlichkeit ist daher eine klare, präzise Darstellungsweise von Sequenzdaten vonnöten.

Schließlich gibt es auch keine vollständigen Datenbanken mit den in Patenten offenbarten Sequenzdaten. Dies bringt eine weitere Schwierigkeit mit sich, da sich viele sehr lange Sequenzen de facto nur mit rechnergestützten Verfahren rationell recherchieren lassen. Probleme treten auch bei der Veröffentlichung auf. Im Rahmen der derzeitigen Druckverfahren werden die Sequenzen einfach als "embedded images" eingefügt, um eine nochmalige Erfassung wegen des damit verbundenen hohen Fehlerrisikos zu vermeiden.

Die angesprochenen Probleme haben das EPA veranlaßt, zu sondieren, wie die Veröffentlichung und die Recherche auf dem Gebiet der Nuclein- und Aminosäuresequenzen mit Hilfe moderner technischer Mittel verbessert werden könnten. Das EPA hat in dieser Sache zunächst Gespräche mit Vertretern der nationalen Ämter und schließlich eingehende Diskussionen mit dem Patent- und Markenamt der Vereinigten Staaten (USPTO) und dem japanischen Patentamt (JPO) geführt, deren Ziel es war,

- ein standardisiertes Format für die Darstellung der Sequenzen festzulegen und

- das weitere Vorgehen der drei Ämter bei der Erfassung und dem Austausch von Sequenzdaten in elektronischer Form im Hinblick auf ihre Verwendbarkeit für Computerrecherchen abzustimmen.

Die Verwendung eines Standardformats bietet nicht nur den Patentämtern und den Anmeldern, die das neue Verfahren anwenden, sondern auch der Öffentlichkeit eine Reihe von Vorteilen:

ised symbols (1-letter or 3-letter code for amino acids), or the use of single- or double-stranded DNA, or the corresponding RNA, arbitrary numbering, or vague and ambiguous indications of manipulated sequence parts. This lack of consistent symbol use and of a standard format for the sequences adds to the difficulties encountered in search and analysis and results in a need for the clear, precise representation of sequence material for the benefit of the examiners and for the public.

Finally, there are no complete databases containing sequence data disclosed in patents. Since many very long sequences can in fact only be searched effectively using computer techniques, this poses an added difficulty. Problems are also encountered in the publication of the sequences. Under existing printing methods they are simply inserted as embedded images to avoid the need for rekeying which entails a high risk of errors.

These problems led the EPO to start looking for ways of using new technology to improve publication and search in the field of nucleic acid and amino acid sequences. After contact had been made with representatives of the national Offices, detailed discussions took place with the United States Patent and Trademark Office (USPTO) and with the Japanese Patent Office (JPO) with a view to:

- standardising the representation format of the sequences

- co-ordinating, between the three Offices, further action relating to the capture and exchange of sequence data in electronic form in order to make them computer-searchable.

The use of a standard format offers a number of advantages to patent offices and applicants adopting the new method, as well as to the public:

de diverses manières : symboles non normalisés (code à une ou à trois lettres pour les acides aminés), utilisation de l'ADN simple ou double brin ou de l'ARN correspondant, numérotation arbitraire, indications vagues et ambiguës concernant les régions manipulées de la séquence. Ce manque de cohérence dans l'utilisation des symboles et l'absence de format normalisé pour les séquences ajoutent aux difficultés rencontrées lors de la recherche et de l'analyse, et rendent nécessaire une représentation claire et précise du matériel séquentiel, dans l'intérêt des examinateurs comme du public.

Enfin, il n'existe pas de base de données complètes contenant des données relatives à des séquences divulguées dans des brevets, ce qui constitue une difficulté supplémentaire étant donné qu'un grand nombre de très longues séquences ne peuvent faire l'objet d'une recherche efficace qu'à l'aide de l'informatique. La publication des séquences pose également problème. Dans les techniques d'impression actuelles, les séquences sont simplement insérées comme image dans le corps du texte afin d'éviter une nouvelle saisie au clavier, qui comporte un grand risque d'erreur.

Ces problèmes ont amené l'OEB à rechercher des moyens permettant d'améliorer la publication et la recherche concernant des séquences d'acides nucléiques et d'acides aminés en recourant à des techniques nouvelles. Après de premiers contacts avec les représentants des offices nationaux, des discussions approfondies ont été engagées avec l'Office des brevets et des marques des Etats-Unis (USPTO) et avec l'Office japonais des brevets (JPO) en vue de :

- normaliser le format de représentation des séquences,

- coordonner entre les trois offices toute action concernant la saisie et l'échange des données relatives aux séquences sous forme électronique, afin d'en permettre une recherche informatisée.

L'utilisation d'un format normalisé offre un certain nombre d'avantages non seulement aux offices de brevets et aux demandeurs adoptant la nouvelle procédure, mais aussi au public :

- Die Anmelder werden bei den Schreibarbeiten entlastet, da die Vorteile der Daten- und Textverarbeitung genutzt werden können.

- Bei der Digitalisierung der Anmeldungen für die Veröffentlichung können Kosten eingespart werden.

- Beim Druck der Sequenzen in den Patentanmeldungen und Patenten wird die Fehlerquote verringert und die Qualität verbessert.

- Die Sequenzen können von den Anmeldern, der Öffentlichkeit und den Prüfern leichter ausgewertet werden; dadurch erhöht sich die Prüfungseffizienz und die Rechtssicherheit.

- Die Daten können in elektronischer Form besser zugänglich gemacht und weitergegeben werden, da die vorhandenen Datenbanken für Nichtpatentliteratur durch eine neue patentspezifische Sequenzdatenbank ergänzt werden.

- Es wird ein Zugriff auf die mit dem USPTO und dem JPO ausgetauschten Sequenzen (dreiseitiges Projekt zum Austausch von Sequenzen) ermöglicht.

- Auf dem Gebiet der Biotechnologie wird die Verwendung allgemein anerkannter Standards gefördert.

## 2. Standardisierung

Ausgehend von den vorstehenden Überlegungen hat das EPA ein Standardformat für die Darstellung von Nucleotid- und Aminosäuresequenzen festgelegt; dies ist in Absprache mit dem USPTO und dem JPO geschehen, die dasselbe Format verwenden werden.

Ab 1. April 1990 wird nachdrücklich empfohlen, in europäischen Patentanmeldungen mit Nucleotid- und/oder Aminosäuresequenzen, die unter die vorgegebenen Definitionen fallen, die in Anlage 1 festgelegten Standardsymbole und Darstellungsweisen zu verwenden. Später sollen anstelle dieser Empfehlung entsprechende zwingende Vorschriften in die Ausführungsordnung zum EPÜ aufgenommen werden.

Ab 1. Januar 1990 wird auch das JPO Patentanmeldungen im selben Standardformat und mit denselben Symbolen annehmen, während das USPTO die Verwendung dieses Standardformats von sich aus empfiehlt, wobei die entspre-

- it means less paperwork for the applicant, enabling the advantages of data and word processing to be utilised

- it reduces the cost of converting applications into digitised form for publication

- it enhances the accuracy and quality of the printed sequences in patent applications and patents

- it makes for easier interpretation of sequences by applicants, the public and examiners alike, resulting in more efficient examination and improved legal certainty

- it will improve access to and dissemination of the data in electronic form because the existing non-patent literature databases will be complemented by a new patent sequence database

- it will open up access to the sequence data exchanged with the USPTO and JPO in a Trilateral Sequence Exchange Project

- it will encourage the use of generally accepted standards in the biotechnology world.

## 2. Standardisation

In view of the above, the EPO has established a standard representation for nucleotide and amino acid sequences in accordance with the agreements with the USPTO and the JPO, who will use the same format.

From 1 April 1990 applicants will be strongly recommended to use the standard symbols and representation methods set out in Annex 1 in European patent applications containing nucleotide and/or amino acid sequences falling within the definitions used. The recommendation will eventually be replaced by mandatory provisions in the EPC Implementing Regulations.

From 1 January 1990, patent applications using the same standard format and symbols will be accepted by the JPO and their use will be recommended by the

- réduction des travaux sur papier pour le demandeur, puisqu'il est possible de bénéficier des avantages offerts par l'informatique et le traitement de textes;

- réduction du coût de la conversion des demandes sous une forme numérisée en vue de leur publication ;

- amélioration de la précision et de la qualité des séquences imprimées dans les demandes et les brevets ;

- compréhension des séquences facilitée pour les demandeurs, le public et les examinateurs, d'où un examen plus efficace et une plus grande sécurité juridique;

- amélioration de l'accès aux données informatisées et de leur diffusion grâce à la création d'une nouvelle base de données relatives aux séquences contenues dans les brevets venant compléter les bases de données non-brevet ;

- accès aux données relatives aux séquences échangées avec l'USPTO et le JPO dans le cadre d'un projet tripartite d'échange de séquences ;

- incitation à utiliser des normes universellement acceptées en biotechnologie.

## 2. Normalisation

Compte tenu de ce qui précède, l'OEB a mis au point une représentation normalisée des séquences nucléotidiques et d'acides aminés conforme aux accords passés avec l'USPTO et le JPO, qui utiliseront le même format.

A compter du 1er avril 1990, il sera vivement recommandé aux demandeurs d'utiliser, dans les demandes de brevet européen traitant de séquences nucléotidiques et/ou d'acides aminés entrant dans les définitions données, les symboles normalisés et les méthodes de représentation figurant à l'Annexe 1. Cette recommandation sera ultérieurement remplacée par des dispositions impératives dans le règlement d'exécution de la CBE.

A compter du 1er janvier 1990, les demandes de brevet utilisant le même format et les mêmes symboles normalisés seront acceptées par le JPO, et l'utilisation de ce format et de ces symboles sera recommandée par l'USPTO, auprès duquel

chenden Daten in maschinenlesbarer Form auf Diskette oder Band eingereicht werden müssen.

Das EPA verlangt die Daten nicht auf bestimmten Datenträgern; die Anmelder können die betreffenden Anmeldungen in gedruckter Form in maschinenlesbarer OCR-Schrift einreichen.

Um die Einhaltung der Vorschriften für die standardisierte Darstellung zu erleichtern, könnte das EPA der Öffentlichkeit, soweit ein entsprechender Bedarf besteht, eine Diskette mit einem Eingabeprogramm (namens AUTHORIN) zur Verfügung stellen, das alle verlangten Angaben nach einem vorgegebenen, standardisierten Schema abfragt. Dieses Programm würde sich so eng wie möglich an das von den sechs großen Datenbankproduzenten für Nichtpatentliteratur verwendete Einreichungsformblatt anlehnen; es soll dann auch beim USPTO erhältlich sein.

### 3. Erläuterung der Regeln

In Anlage 1 zur vorliegenden Mitteilung sind die einschlägigen Regeln für die standardisierte Darstellung von Nucleotid- und Aminosäuresequenzen in europäischen Patentanmeldungen dargelegt.

Unter Nummer 1.1 wird erläutert, für welche Erfindungen die Regeln gelten. Wenn in Patentanmeldungen Nucleotid- oder Aminosäuresequenzen offenbart werden und der Anmelder die Sequenzen selbst darstellen möchte, sollte die Wiedergabe entsprechend den nun festgelegten Standards erfolgen.

Nummer 1.1 enthält ferner die Definition einer Nucleotid- und Aminosäuresequenz und legt fest, ab welcher Zahl von Aminosäuren oder Nucleotiden in einer Sequenz die standardisierte Darstellung zu verwenden ist. Von der standardisierten Darstellung werden dabei Sequenzen ausgeschlossen, deren Nucleotide und Aminosäuren nicht mit den unter den Nummern 2.3 und 2.6 aufgeführten Symbolen wiedergegeben werden können. Aminosäuresequenzen werden dadurch auf Sequenzen aus den gängigen L-Aminosäuren der in der Natur vorkommenden Proteine beschränkt. Mit dieser Definition werden auch das JPO und das USPTO arbeiten.

Unter den Nummern 1.2 bis 1.5 wird das Konzept des "Sequenzprotokolls" vorgestellt und erläutert. Die Angaben zu jeder Sequenz oder Teilsequenz gliedern sich

USPTO, where the relevant data must be submitted in machine-readable form on diskette or tape.

The EPO will not require the data on specific data carriers and applicants may file such applications in printed form using machine-readable OCR characters.

To facilitate compliance with the standard representation rules and if there is sufficient demand, the EPO may make available to the public a diskette with an input program (called AUTHORIN) which elicits all the information required in a prepared, standardised form. This program would follow as closely as possible the submission form used by the six major non-patent literature database producers and will also be distributed by the USPTO.

### 3. Explanation of the rules

Annex 1 to this notice sets out the rules for the standard representation of nucleotide and amino acid sequences in European patent applications.

Paragraph 1.1 explains what inventions are affected by the rules. Patent applications containing disclosures of nucleotide or amino acid sequences in which the applicant wishes to represent the sequence itself should include these representations in conformity with the standards.

Paragraph 1.1 also defines the terms "nucleotide" and "amino acid" sequence stipulating the number of amino acids or nucleotides in a sequence above which the standard representation is to be used. It furthermore specifies that the standard representation should only be used for those sequences whose nucleotides and amino acids can be represented by the symbols listed in paragraphs 2.3 and 2.6. Accordingly, amino acid sequences only include those containing L-amino acids commonly found in naturally occurring proteins. The same definition will be used by the JPO and the USPTO.

Paragraphs 1.2 to 1.5 introduce and explain the concept of "Sequence Listing". For each sequence or part of a sequence the information is divided into

les données devront être déposées sur disquette ou sur bande sous une forme lisible par machine.

L'OEB n'exigera pas que ces données soient fournies sur des supports de données particuliers, et les demandeurs pourront déposer ces demandes sous forme imprimée en utilisant des caractères ROC lisibles par machine.

Afin de faciliter le respect des règles de représentation normalisée et si la demande est suffisante, l'OEB mettra éventuellement à la disposition du public une disquette comportant un programme (appelé AUTHORIN) prévoyant toutes les informations requises sous une forme normalisée. Ce programme sera aussi conforme que possible au formulaire utilisé par les six principaux producteurs de bases de données non-brevet et sera également diffusé par l'USPTO.

### 3. Explication des règles

L'Annexe 1 au présent communiqué contient les règles concernant la représentation normalisée de séquences nucléotidiques et d'acides aminés dans les demandes de brevet européen.

Le point 1.1 mentionne les inventions concernées par les règles. Lorsque le demandeur souhaite représenter une séquence dans une demande de brevet divulguant des séquences nucléotidiques ou d'acides aminés, cette représentation devrait être conforme aux normes prescrites.

Il définit également les termes "séquences nucléotidiques" et/ou "d'acides aminés" et précise le nombre d'acides aminés ou de nucléotides d'une séquence au-dessus duquel la représentation normalisée devra être utilisée. Cette représentation normalisée ne devrait s'appliquer qu'aux séquences dont les nucléotides et acides aminés peuvent être représentés par les symboles énumérés aux points 2.3 et 2.6. Par conséquent, les séquences d'acides aminés seront limitées à celles contenant des acides aminés L que l'on rencontre généralement dans les protéines naturelles. La même définition sera adoptée par le JPO et l'USPTO.

Les points 1.2 à 1.5 présentent et expliquent le concept de "Liste des séquences". Pour chaque séquence ou région d'une séquence, l'information est

in die drei unter den Nummern 1.3 bis 1.5 beschriebenen Teile. Die Einführung eines Sequenzprotokolls als Bestandteil der Beschreibung ist mit dem JPO und dem USPTO vereinbart worden. Somit werden bei allen drei Ämtern im Sequenzprotokoll dieselben - entweder zwingend vorgeschriebenen oder freiwillig eingereichten - Angaben enthalten sein.

Unter Nummer 1.6 wird empfohlen, die Sequenzkennzahlen immer dann zu verwenden, wenn in der Beschreibung oder in den Ansprüchen einer Patentanmeldung von einer Sequenz die Rede ist, auch wenn im Text oder in den Abbildungen der Anmeldung die Sequenz selbst oder weitere oder abgewandelte Darstellungen der Sequenz vorkommen.

Nummer 1.7 gibt Aufschluß über die nach den derzeitigen Erfordernissen bestehenden Einreichungsmöglichkeiten für die einschlägigen Anmeldungen; wie verfahren wird, wenn die Erfordernisse nicht erfüllt sind, ist Nummer 1.8 zu entnehmen.

Unter den Nummern 2.1 bis 2.15 ist dargelegt, in welchem Format und mit welchen Symbolen die Daten der Nucleotid- und/oder Aminosäuresequenzen aufzuführen und wie die Zeichen zu gruppieren, räumlich zu trennen und zu nummerieren sind. Auch diese Formatierungsverfahren sind mit dem JPO und dem USPTO für den elektronischen Datenaustausch vereinbart worden.

Unter den Nummern 3.1 bis 3.10 werden schließlich verschiedene Angaben behandelt, die der Anmelder freiwillig machen kann, um das sequenzierte Segment und die Merkmale der Sequenz genauer zu beschreiben. Im Sequenzprotokoll brauchen nur die auf eine aufgeführte Sequenz zutreffenden Angaben gemacht zu werden.

Sämtliche Definitionen sowie die Standardsymbole und Darstellungsverfahren für Nucleotid- oder Aminosäuresequenzen sind im Rahmen eines dreiseitigen Projekts in enger Zusammenarbeit mit dem USPTO und dem JPO festgelegt worden. Alle drei Ämter werden den Anmeldern somit künftig die Einhaltung derselben Regeln und Definitionen empfehlen oder zwingend vorschreiben und entsprechend wiedergegebene Informationen annehmen.

Ein Muster eines Sequenzprotokolls ist Anlage 2 der vorliegenden Mitteilung zu entnehmen.

three parts described in paragraphs 1.3 to 1.5. The introduction of the Sequence Listing as part of the description has been agreed with the JPO and the USPTO. The various items of information contained in the Sequence Listing, whether compulsory or optional, will be identical for all three Offices.

Paragraph 1.6 recommends the use of the sequence identifiers in all instances where the description or claims of a patent application discuss a sequence, regardless of whether the sequence or other additional or modified representations of the sequence are embedded in the text or in the figures of the application.

Paragraph 1.7 indicates the alternative modes of filing which may be employed for the relevant applications in accordance with current requirements, and paragraph 1.8 lays down the procedure if the requirements are not satisfied.

Paragraphs 2.1 to 2.15 set forth the format and symbols to be used to list the nucleotide and/or amino acid sequence data, as well as the manner in which the characters should be grouped, spaced and numbered. These formatting procedures have also been agreed with the JPO and the USPTO for electronic data exchange.

Finally, paragraphs 3.1 to 3.10 list a number of optional items providing a more detailed description of the sequenced segment and the features of the sequence. Only information applicable to a listed sequence may be given in the Sequence Listing.

All the definitions and the set of standard symbols and representation methods for nucleotide or amino acid sequences have been drawn up in close co-operation with the USPTO and the JPO as a Trilateral Project. All three Offices will recommend or require applicants to observe the same set of rules and definitions and accept information presented in conformity with them.

A specimen Sequence Listing is given in Annex 2 to this notice.

divisée en trois parties définies aux points 1.3 à 1.5. L'incorporation de la liste des séquences dans la description a été décidée en accord avec le JPO et l'USPTO. Les différents éléments d'information contenus dans la liste des séquences, qu'ils soient obligatoires ou facultatifs, seront identiques pour les trois offices.

Le point 1.6 recommande l'utilisation des numéros d'identification attribués aux séquences dans tous les cas où la description ou les revendications d'une demande de brevet portent sur une séquence, même si la séquence ou des représentations supplémentaires ou modifiées de la séquence sont comprises dans le texte ou dans les figures de la demande.

Le point 1.7 indique quels sont les divers modes de dépôt répondant aux conditions actuelles ; le point 1.8 décrit la procédure à suivre si ces conditions ne sont pas remplies.

Les points 2.1 à 2.15 présentent le format et les symboles à utiliser pour établir la liste des données relatives aux séquences nucléotidiques et/ou d'acides aminés ainsi que la manière dont les caractères doivent être groupés, espacés et numérotés. Ces modes de formatage ont également été convenus avec le JPO et l'USPTO pour l'échange de données électroniques.

Enfin, les points 3.1 à 3.10 énumèrent une série d'indications facultatives permettant de fournir une description plus détaillée du segment séquencé et des caractéristiques de la séquence. Seules les informations concernant une séquence figurant sur la liste peuvent figurer sur la liste des séquences.

Toutes les définitions ainsi que les symboles normalisés et les modes de représentation de séquences nucléotidiques ou d'acides aminés ont été élaborés en étroite collaboration avec l'USPTO et le JPO dans le cadre d'un projet tripartite. Les trois offices recommanderont aux demandeurs ou exigeront d'eux qu'ils s'en tiennent aux mêmes règles et définitions et accepteront les données qui leur seront conformes.

Un exemple de liste des séquences figure à l'Annexe 2 au présent communiqué.

**ANLAGE 1****Regeln für die standardisierte Darstellung von Nucleotid- und Aminosäuresequenzen in Patentanmeldungen****1. Definitionen, Geltungsbereich und Verfahren**

1.1 Betrifft oder umfaßt eine Erfindung Sequenzen von mindestens 10 Nucleotiden (Nucleotidsequenzen) oder unverzweigte Sequenzen von mindestens 4 Aminosäuren (Aminosäuresequenzen), so ist jede in einer Anmeldung gesondert offenbarte Sequenz entsprechend den nachstehend beschriebenen Standards darzustellen. Verzweigte Sequenzen sind aus dieser Definition ausdrücklich ausgeschlossen.

Im Rahmen der vorliegenden Regeln

- sind unter Nucleotiden nur die Nucleotide zu verstehen, die mittels der unter Nummer 2.3 aufgeführten Symbole wiedergegeben werden können. Modifizierte Basen können nach der Anleitung unter Nummer 2.4 beschrieben werden, sind aber in der Nucleotidsequenz nicht explizit auszuweisen;

- sind unter Aminosäuren die unter Nummer 2.6 aufgeführten gängigen L-Aminosäuren aus den in der Natur vorkommenden Proteinen zu verstehen. Posttranslational modifizierte Aminosäuren können nach der Anleitung unter Nummer 2.7 beschrieben werden, sind aber in der Aminosäuresequenz nicht explizit auszuweisen.

Unter diese Definition fallen auch Peptide oder Proteine, die anhand der unter Nummer 2.6 aufgeführten Symbole sowie einer an anderer Stelle in das Sequenzprotokoll aufgenommenen Beschreibung, die beispielsweise Aufschluß über ungewöhnliche Bindungen, Quervernetzungen und "end caps", Nichtpeptidbindungen usw. gibt, als Sequenz wiedergegeben werden können.

1.2 Die Sequenzen sind in einem gesonderten, mit "Sequenzprotokoll" überschriebenen Teil der Beschreibung unmittelbar vor den Ansprüchen anzugeben.

**ANNEX 1****Rules for the standard representation of nucleotide and amino acid sequences In patent applications****1. Definitions, scope and procedure**

1.1 If an invention concerns or involves sequences of ten or more nucleotides (nucleotide sequence) or unbranched sequences of four or more amino acids (amino acid sequence) the representation of each sequence separately disclosed in the application should conform to the standards described below. Branched sequences are specifically excluded from this definition.

For the purposes of these rules:

- "nucleotide" means only those nucleotides that can be represented using the symbols listed in paragraph 2.3. Modified bases may be described as set forth in paragraph 2.4 but should not be shown explicitly in the nucleotide sequence.

- "amino acid" means those L-amino acids commonly found in naturally occurring proteins and listed in paragraph 2.6. Post-translationally modified amino acids may be described as set forth in paragraph 2.7 but should not be shown explicitly in the amino acid sequence.

Any peptide or protein that can be expressed as a sequence using the symbols listed in paragraph 2.6 together with a description elsewhere in the Sequence Listing to describe, for example, abnormal linkages, cross links and end caps, non-peptidyl bonds, etc. is embraced by this definition.

1.2 The sequences should be filed in a separate part of the description entitled "Sequence Listing," immediately prior to the claims.

**ANNEXE 1****Règles concernant la représentation normalisée de séquences nucléotidiques et d'acides aminés dans les demandes de brevet****1. Définitions, champ d'application des règles et procédure à suivre**

1.1 Si une invention concerne ou implique des séquences de dix nucléotides ou plus (séquence nucléotidique) ou des séquences linéaires de quatre acides aminés ou plus (séquence d'acides aminés), la représentation de chaque séquence exposée séparément dans la demande devra être conforme aux normes suivantes. Les séquences non linéaires sont expressément exclues de cette définition.

Aux fins des présentes règles, on entend par:

- "nucléotide" seulement les nucléotides pouvant être représentés à l'aide des symboles figurant au point 2.3. Les bases modifiées peuvent être décrites comme indiqué au point 2.4 ; elles ne devraient cependant pas être représentées de façon explicite dans la séquence nucléotidique;

- "acide aminé" les acides aminés L que l'on rencontre généralement dans des protéines naturelles et visés au point 2.6. Les acides aminés modifiés après traduction peuvent être décrits comme indiqué au point 2.7 ; ils ne devraient cependant pas être représentés de façon explicite dans la séquence d'acides aminés.

Entre dans cette définition tout peptide ou protéine pouvant être exprimé par une séquence à l'aide des symboles énumérés au point 2.6 et accompagné par exemple, dans la liste des séquences, d'une description des liaisons anormales, des liaisons croisées et des coiffes terminales, des liaisons non peptidiques, etc.

1.2 Les séquences devraient figurer dans une partie distincte de la description, intitulée "Liste des séquences" et précédant immédiatement les revendications.

1.3 Das Sequenzprotokoll muß **zu Jeder Sequenz oder Teilsequenz Jeweils** folgende Angaben in nachstehender Reihenfolge enthalten:

- eine Sequenzkennzahl in der vereinfachten Schreibweise

SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 usw.

- die Sequenzlänge, ausgedrückt als Anzahl der Basenpaare oder Aminosäuren

- die Art der Sequenz, d. h. Nucleotid- oder Aminosäuresequenz.

1.4 Im Anschluß hieran können die nachstehend unter den Nummern 3.1 bis 3.10 aufgeführten näheren Angaben gemacht werden.

1.5 Danach ist die eigentliche Sequenz oder Teilsequenz, die an einer beliebigen Stelle der Anmeldung offenbart sein kann, gemäß der Anleitung unter den Nummern 2.1 bis 2.15 wiederzugeben.

1.6 In der Beschreibung, den Ansprüchen oder der Zusammenfassung der Anmeldung ist auf die im Sequenzprotokoll dargestellten Sequenzen anhand der Kennzahl zu verweisen, auch wenn die Sequenz selbst oder weitere oder abgewandelte Darstellungen der Sequenz im Text oder in den Abbildungen enthalten sind.

1.7 **Ab 1. April 1990** sind Anmeldungen mit Sequenzprotokoll entsprechend den vorstehenden Regeln:

- entweder in gedruckter Form in Standard-OCR-Schrift

- oder zusammen mit einer vom EPA bereitgestellten Diskette (mit dem Sequenzeingabeprogramm AUTHORIN), auf der das Sequenzprotokoll aufgezeichnet ist, einzureichen.

1.8 Sind die Erfordernisse des Sequenzprotokolls bei Einreichung der Anmeldung nicht erfüllt, so teilt die Eingangsstelle dies dem Anmelder mit und fordert ihn auf, das Sequenzprotokoll nachzureichen. Jedem daraufhin eingereichten Sequenzprotokoll ist eine Erklärung beizufügen, daß das Protokoll keine neuen Angaben enthält.

1.3 The Sequence Listing should include, **for every sequence or part of a sequence separately**, and in the following order:

- a sequence identifier number, for convenient reference, written as

SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, etc.;

- the length of the sequence in number of base pairs or amino acids.

- the type of sequence, i.e. nucleotides or amino acids.

1.4 This information may be followed by the particulars set out in paragraphs 3.1 to 3.10 hereinbelow.

1.5 The sequence itself or part of a sequence disclosed in any part of the application, represented as described in paragraphs 2.1 to 2.15, should then be listed.

1.6 In the description, claims or abstract of the application, the sequences represented in the Sequence Listing should be referred to by their identifier number, even if the sequence or other additional or modified representations of the sequence are embedded in the text or in the figures.

1.7 As from **1 April 1990** applications containing Sequence Listings should be filed in conformity with the above rules:

- in printed form only in standard OCR characters, or

- together with a diskette distributed by the EPO (containing a sequence input program called AUTHORIN) on which the Sequence Listing has been entered.

1.8 If the Sequence Listing requirements are not satisfied at the time of filing, the Receiving Section will inform the applicant accordingly and invite him to file the Sequence Listing. Any Sequence Listing submitted in response should be accompanied by a statement that it includes no new matter.

1.3 La liste des séquences devrait comporter pour **chaque séquence ou région d'une séquence**, et dans l'ordre suivant:

- un numéro d'identification de séquence destiné à faciliter les références et qui s'écrira

SEQ ID NO : 1, SEQ ID NO : 2, SEQ ID NO : 3, etc.

- la longueur de la séquence exprimée en nombre de paires de bases ou d'acides aminés

- le type de séquence, c'est-à-dire nucléotides ou acides aminés.

1.4 Ces indications peuvent être complétées par les renseignements visés aux points 3.1 à 3.10 ci-après.

1.5 Enfin, la séquence elle-même ou une région d'une séquence décrite dans la demande et représentée conformément aux points 2.1 à 2.15 devrait être portée sur la liste.

1.6 Dans la description, les revendications ou l'abrégé de la demande, il devrait être fait référence aux séquences figurant dans la liste des séquences en utilisant leur numéro d'identification, même si des représentations supplémentaires ou modifiées de la séquence sont comprises dans le texte ou dans les figures.

1.7 A compter du **1er avril 1990**, toute demande contenant des listes de séquences devrait être déposée conformément aux règles susmentionnées:

- sous forme imprimée, uniquement en caractères ROC normalisés

- ou accompagnée d'une disquette fournie par l'OEB (et contenant un programme appelé AUTHORIN) sur laquelle la liste des séquences a été chargée.

1.8 Si les conditions relatives à la liste des séquences ne sont pas remplies à la date de dépôt, la section de dépôt le signale au demandeur et l'invite à déposer cette liste. Toute liste de séquences adressée en réponse à cette invitation devrait être accompagnée d'une déclaration selon laquelle la liste présentée ne comporte aucun élément nouveau.

## 2. Standardformat für Sequenzdaten

Nucleotid- und Aminosäuresequenzen sind in folgendem Format wiederzugeben:

2.1 Für die Darstellung der Nucleotid- und Polypeptidsequenzen kommt eine dreidreifolgende Möglichkeiten in Frage:

- bloße Nucleotidsequenz, wenn keine entsprechende codierende Sequenz (CDS) angegeben ist oder es sich um eine nichtcodierende Sequenz handelt
- Polypeptidsequenz, wenn keine entsprechende Polynucleotidsequenz angegeben ist
- Nucleotidsequenz zusammen mit dem entsprechenden Protein

2.2 **Nucleotidsequenzen** sind nur anhand eines Einzelstrangs in 5'-3'-Richtung von links nach rechts wiederzugeben.

2.3 Die **Basen** einer Nucleotidsequenz sind anhand des einbuchstabigen Codes für Nucleotidsequenzzeichen entsprechend der folgenden Liste (IUPAC-IUB-Standards, s. Nucleic Acids Research 13, 3021 - 3030 (1985) und The Biochemical Journal 219, Nr. 2, 345 - 373 (1984)) darzustellen:

Symbol	Bedeutung	Ableitung der Bezeichnung
A	A	Adenin
G	G	Guanin
C	C	Cytosin
T	T	Thymin
U	U	Uridin
R	G oder A	PuRin
Y	T/U oder C	PYrimidin
M	A oder C	AMino
K	G oder T/U	Keto
S	G oder C	Starke Bindungen 3 H-Brücken
W	A oder T/U	schwache (E: Weak) Bindungen 2 H-Brücken
B	G oder C oder T/U	nicht A
D	A oder G oder T/U	nicht C

## 2. Standard format of sequence data

Nucleotide and amino acid sequences should be represented in the following format:

2.1 Nucleotide and polypeptide sequences should be represented by one of the following three possibilities:

- a pure nucleotide sequence, in those cases where no corresponding CDS (coding sequence) is given or for non-coding sequences
- a polypeptide sequence, in those cases where no corresponding polynucleotide sequence is given
- a nucleotide sequence together with its corresponding protein.

2.2 **A nucleotide sequence** should be presented only by a single strand, in the 5' to 3' direction from left to right.

2.3 The **bases** of a nucleotide sequence should be represented using the one-letter code for nucleotide sequence characters in conformity with the following list (IUPAC-IUB standards described in Nucleic Acids Research 13, 3021-3030 (1985) and in The Biochemical Journal 219, No. 2, 345-373 (1984):

Symbol	Meaning	Origin of designation
A	A	Adenine
G	G	Guanine
C	C	Cytosine
T	T	Thymine
U	U	Uridine
R	G or A	puRine
Y	T/U or C	pYrimidine
M	A or C	aMino
K	G or T/U	Keto
S	G or C	Strong interactions 3 H-bonds
W	A or T/U	Weak interactions 2 H-bonds
B	G or C or T/U	not A
D	A or G or T/U	not C

## 2. Format normalisé des données relatives aux séquences

Les séquences nucléotidiques et d'acides aminés devraient être représentées dans le format suivant :

2.1 Les séquences nucléotidiques et polypeptidiques devraient être représentées sous l'une des trois formes suivantes :

- une séquence nucléotidique pure en l'absence d'indication d'une séquence codante correspondante ou pour des séquences non codantes ;
- une séquence polypeptidique lorsqu'aucune séquence polynucléotidique correspondante n'est indiquée ;
- une séquence nucléotidique ainsi que la protéine correspondante.

2.2 Une **séquence nucléotidique** devrait être uniquement représentée par un seul brin de codage, dans le sens 5' - 3' et de gauche à droite.

2.3 Les **bases** d'une séquence nucléotidique devraient être représentées au moyen du code à une lettre utilisé pour les séquences nucléotidiques, conformément à la liste suivante (normes IUPAC-IUB figurant dans Nucleic Acids Research 13, 3021-3030 (1985) et dans The Biochemical Journal 219, n°2, 345-373 (1984):

Symbole	Signification	Origine de la désignation
A	A	Adénine
G	G	Guanine
C	C	Cytosine
T	T	Thymine
U	U	Uridine
R	G ou A	PuRine
Y	T/U ou C	pYrimidine
M	A ou C	aMino
K	G ou T/U	Keto (ceto)
S	G ou C	Interactions fortes (3 liaisons d'hydrogène)
W	A ou T/U	Interactions faibles (2 liaisons d'hydrogène)
B	G ou C ou T/U	autre que A
D	A ou G ou T/U	autre que C



H	A oder C oder T/U	nicht G	H	A or C or T/U	not G	H	A ou C ou T/U	autre que G
V	A oder G oder C	nicht T, nicht U	V	A or G or C	not T, not U	V	A ou G ou C	autre que T et U N
N	(A oder G oder C oder T/U) oder (unbekannt oder sonstige)	beliebig (E:aNy)	N	(A or G or C or T/U) or (unknown or other)	aNy	N	(A ou G ou C ou T/U) (n'importe lequel) ou (non connu ou autre)	aNy

2.4 **Modifizierte Basen** sind in der Sequenz selbst wie die entsprechenden unmodifizierten Basen wiederzugeben, wenn es sich um eine der nachstehend aufgeführten modifizierten Basen handelt und die Modifikation an anderer Stelle des Sequenzprotokolls näher beschrieben wird, wobei die nachstehend aufgeführten Codes zu verwenden sind.

Ansonsten sind modifizierte Basen in der Sequenz mit "N" aufzuführen und an anderer Stelle des Sequenzprotokolls nähere Angaben zu machen.

2.4 **Modified bases** should be represented as the corresponding unmodified bases in the sequence itself, if the modification is further described elsewhere in the Sequence Listing, using the codes from the list below.

Otherwise, modified bases should be listed in the sequence as "N", with further information given elsewhere in the Sequence Listing.

2.4 Les **bases modifiées** devraient être représentées par les bases non modifiées correspondantes contenues dans la séquence elle-même si la base modifiée figure parmi celles énumérées ci-après et que la modification fasse l'objet d'une description plus détaillée dans la liste des séquences au moyen des codes contenus dans la liste ci-après.

Sinon, les bases modifiées devraient être désignées par "N" dans la séquence et assorties de renseignements complémentaires dans la liste des séquences.

Symbol	Bedeutung	Symbol	Meaning	Symbole	Signification
ac4c	4-Acetylcytidin	ac4c	4-acetylcytidine	ac4c	4-acétylcytidine
chm5u	5-(Carboxyhydroxymethyl)uridin	chm5u	5-(carboxyhydroxymethyl)uridine	chm5u	5-(carboxyhydroxyméthyl)uridine
cm	2'-O-Methylcytidin	cm	2'-O-methylcytidine	cm	2'-O-méthylcytidine
cmnm5s2u	5-Carboxymethylaminomethyl-2-thiouridin	cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine	cmnm5s2u	5-carboxyméthylaminométhyl-2-thiouridine
cmnm5u	5-Carboxymethylaminomethyluridin	cmnm5u	5-carboxymethylaminomethyluridine	cmnm5u	5-carboxyméthylaminométhyluridine
d	Dihydrouridin	d	dihydrouridine	d	dihydro-uridine
fm	2'-O-Methylpseudo-uridin	fm	2'-O-methylpseudo-uridine	fm	2'-O-méthylpseudo-uridine
gal q	beta, D-Galactosyl-queuosin	gal q	beta, D-galactosyl-queuosine	gal q	bêta, D-galactosyl-queuosine
gm	2'-O-Methylguanosin	gm	2'-O-methylguanosine	gm	2'-O-méthylguanosine
i	Inosin	i	inosine	i	inosine
i6a	N6-Isopentenyladenosin	i6a	N6-isopentenyladenosine	i6a	N6-isopentenyladénosine
m1a	1-Methyladenosin	m1a	1-methyladenosine	m1a	1-méthyladénosine
m1f	1-Methylpseudo-uridin	m1f	1-methylpseudo-uridine	m1f	1-méthylpseudo-uridine
m1g	1-Methylguanosin	m1g	1-methylguanosine	m1g	1-méthylguanosine
m1i	1-Methylinosin	m1i	1-methylinosine	m1i	1-méthylinosine
m22g	2,2-Dimethylguanosin	m22g	2,2-dimethylguanosine	m22g	2,2-diméthylguanosine
m2a	2-Methyladenosin	m2a	2-methyladenosine	m2a	2-méthyladénosine
m2g	2-Methylguanosin	m2g	2-methylguanosine	m2g	2-méthylguanosine
m3c	3-Methylcytidin	m3c	3-methylcytidine	m3c	3-méthylcytidine
m5c	5-Methylcytidin	m5c	5-methylcytidine	m5c	5-méthylcytidine
m6a	N6-Methyladenosin	m6a	N6-methyladenosine	m6a	N6-méthyladénosine
m7g	7-Methylguanosin	m7g	7-methylguanosine	m7g	7-méthylguanosine
mam5u	5-Methylaminomethyluridin	mam5u	5-methylaminomethyluridine	mam5u	5-méthylaminométhyluridine
mam5s2u	5-Methoxyaminomethyl-2-thiouridin	mam5s2u	5-methoxyaminomethyl-2-thiouridine	mam5s2u	5-méthoxyaminométhyl-2-thiouridine
man q	beta, D-Mannose-queuosin	man q	beta, D-mannosyl-queuosine	man q	bêta, D-mannosyl-queuosine

mcm5s2u	5-Methoxycarbonyl-methyl-2-thiouridin	mcm5s2u	5-methoxycarbonyl-methyl-2-thiouridine	mcm5s2u	5-méthoxycarbonyl-méthyl-2-thio-uridine
mcm5u	5-Methoxycarbonyl-methyluridin	mcm5u	5-methoxycarbonyl-methyluridine	mcm5u	5-méthoxycarbonyl-méthyluridine
mo5u	5-Methoxyuridin	mo5u	5-methoxyuridine	mo5u	5-méthoxyuridine
ms2i6a	2-Methylthio-N6-isopentenyladenosin	ms2i6a	2-methylthio-N6-isopentenyladenosine	ms2i6a	2-méthylthio-N6-isopentenyladénosine
ms2t6a	N-((9-beta-D-Ribofuranosyl-2-methylthiopurin-6-yl)carbamoyl)-threonin	ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine	ms2t6a	N-(9-bêta-D-ribofuranosyl-2-méthylthiopurine-6-yl)carbamoyl-thréonine
mt6a	N-((9-beta-D-Ribofuranosylpurin-6-yl)N-methylcarbamoyl)-threonin	mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methylcarbamoyl)threonine	mt6a	N-(9-bêta-D-ribofuranosylpurine-6-yl)N-méthylcarbamoyl-thréonine
mv	Uridin-5-oxyyessigsäuremethylester	mv	uridine-5-oxycetic acid-methylester	mv	ester méthylé d'uridine 5 oxyacétique
o5u	Uridin-5-oxyyessigsäure (v)	o5u	uridine-5-oxycetic acid (v)	o5u	acide d'uridine 5 oxyacétique
osyw	Wybutoxosin	osyw	wybutoxosine	osyw	wybutoxosine
p	Pseudouridin	p	pseudouridine	p	pseudo-uridine
q	Queuosin	q	queuosine	q	queuosine
s2c	2-Thiocytidin	s2c	2-thiocytidine	s2c	2-thiocytidine
s2t	5-Methyl-2-thiouridin	s2t	5-methyl-2-thiouridine	s2t	5-méthyl-2-thio-uridine
s2u	2-Thiouridin	s2u	2-thiouridine	s2u	2-thio-uridine
s4u	4-Thiouridin	s4u	4-thiouridine	s4u	4-thio-uridine
t	5-Methyluridin	t	5-methyluridine	t	5-méthyluridine
t6a	N-((9-beta-D-Ribofuranosylpurin-6-yl)-carbamoyl)threonin	t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)-carbamoyl)threonine	t6a	N-(9-bêta-D-ribofuranosylpurine-6-yl)-carbamoyl-thréonine
tm	2'-O-Methyl-5-methyluridin	tm	2'-O-methyl-5-methyluridine	tm	2'-O-méthyl-5-méthyluridine
um	2'-O-Methyluridin	um	2'-O-methyluridine	um	2'-O-méthyluridine
yw	Wybutosin	yw	wybutosine	yw	wybutosine
x	3-(3-Amino-3-carboxypropyl)-uridin, (acp3)u	x	3-(3-amino-3-carboxy-propyl)-uridine, (acp3)u	x	3-(3-amino-3-carboxypropyl)uridine, (acp3)u

## SONSTIGE

2.5 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.

2.6 Die **Aminosäuren** sind anhand des dreibuchstabigen Codes mit großem Anfangsbuchstaben entsprechend der nachstehenden Liste (IUPAC-IUB-Standards, s. Nucleic Acids Research 13, 3021 - 3030 (1985) und The Biochemical Journal 219, Nr. 2, 345 - 373 (1984)) darzustellen:

Symbol	Bedeutung
Ala	Alanin
Cys	Cystein
Asp	Asparaginsäure
Glu	Glutaminsäure
Phe	Phenylalanin

## OTHER

2.5 The **amino acids** in a protein or peptide sequence should be listed in the amino to carboxy direction from left to right, and the amino and carboxy groups should not be represented in the sequence.

2.6 The **amino acids** should be represented using the three-letter code with the first letter as a capital and should conform to the following list (IUPAC-IUB standards described in Nucleic Acids Research 13, 3021-3030 (1985) and in The Biochemical Journal 219, No 2, 345-373 (1984):

Symbol	Meaning
Ala	Alanine
Cys	Cysteine
Asp	Aspartic Acid
Glu	Glutamic Acid
Phe	Phenylalanine

## AUTRES

2.5 Les **acides aminés** dans une séquence de protéines ou de peptides devraient être mentionnés dans la liste dans le sens amino-carboxy et de gauche à droite, alors que les groupes amino et carboxyliques ne devront pas être représentés dans la séquence.

2.6 Les **acides aminés** devraient être représentés au moyen du code à trois lettres, la première lettre étant en majuscule, conformément à la liste suivante (normes IUPAC-IUB figurant dans Nucleic Acids Research 13, 3021-3030 (1985) et dans The Biochemical Journal 219, N 2, 345-373 (1984):

Symbole	Signification
Ala	Alanine
Cys	Cystéine
Asp	Acide aspartique
Glu	Acide glutamique
Phe	Phénylalanine

Gly	Glycin	Gly	Glycine	Gly	Glycine
His	Histidin	His	Histidine	His	Histidine
Ile	Isoleucin	Ile	Isoleucine	Ile	Isoleucine
Lys	Lysin	Lys	Lysine	Lys	Lysine
Leu	Leucin	Leu	Leucine	Leu	Leucine
Met	Methionin	Met	Methionine	Met	Méthionine
Asn	Asparagin	Asn	Asparagine	Asn	Asparagine
Pro	Prolin	Pro	Proline	Pro	Proline
Gln	Glutamin	Gln	Glutamine	Gln	Glutamine
Arg	Arginin	Arg	Arginine	Arg	Arginine
Ser	Serin	Ser	Serine	Ser	Sérine
Thr	Threonin	Thr	Threonine	Thr	Thréonine
Val	Valin	Val	Valine	Val	Valine
Trp	Tryptophan	Trp	Tryptophan	Trp	Tryptophane
Tyr	Tyrosin	Tyr	Tyrosine	Tyr	Tyrosine
Asx	Asp oder Asn	Asx	Asp or Asn	Asx	Asp ou Asn
Glx	Glu oder Gln	Glx	Glu or Gln	Glx	Glu ou Gln
Xaa	unbekannt oder sonstige	Xaa	unknown or other	Xaa	indéterminé ou autre

**2.7 Modifizierte und seltene Aminosäuren** sind in der Sequenz selbst wie die entsprechenden unmodifizierten Aminosäuren wiederzugeben, wenn es sich um eine der nachstehend aufgeführten modifizierten Aminosäuren handelt und die Modifikation an anderer Stelle des Sequenzprotokolls näher beschrieben wird, wobei die nachstehend aufgeführten Codes zu verwenden sind.

Ansonsten sind modifizierte oder seltene Aminosäuren in der Sequenz mit "Xaa" aufzuführen und an anderer Stelle des Sequenzprotokolls nähere Angaben zu machen.

**2.7 Modified and unusual amino acids** should be represented as the corresponding unmodified amino acids in the sequence itself if the modified amino acid is one of those listed below and the modification is also further described elsewhere in the Sequence Listing, using the codes from the list below.

Otherwise, modified or unusual amino acids should be listed in the sequence as "Xaa", with further information given elsewhere in the Sequence Listing.

**2.7 Les acides aminés modifiés et peu connus** devraient être représentés par les acides aminés non modifiés correspondants dans la séquence elle-même si l'acide aminé modifié figure parmi ceux énumérés ci-après et que la modification fasse également l'objet d'une description plus détaillée dans la liste des séquences au moyen des codes contenus dans la liste ci-après.

Sinon, les acides aminés modifiés ou peu connus devraient être désignés par "Xaa" dans la séquence et assortis de renseignements complémentaires dans la liste des séquences.

Symbol	Bedeutung	Symbol	Meaning	Symbole	Signification
Aad	2-Aminoadipinsäure	Aad	2-Aminoadipic acid	Aad	acide 2-amino-adipique
bAad	3-Aminoadipinsäure	bAad	3-Aminoadipic acid	bAad	acide 3-amino-adipique
bA1 a	beta-Alanin,#beta-Aminopropionsäure	bA1 a	beta-Alanine,#beta-Amino-propionic acid	bA1a	bêta-alanine, acide bêta-amino-propionique
Abu	2-Aminobuttersäure	Abu	2-Aminobutyric acid	Abu	acide 2-amino-butyrique
4Abu	4-Aminobuttersäure, Piperidinsäure	4Abu	4-Aminobutyric acid, piperidinic acid	4Abu	acide 4-amino-butyrique, acide pipéridinique
Acp	6-Aminocapronsäure	Acp	6-Aminocaproic acid	Acp	acide 6-amino-caproïque
Ahe	2-Aminoheptansäure	Ahe	2-Aminoheptanoic acid	Ahe	acide 2-amino-heptanoïque
Aib	2-Aminoisobuttersäure	Aib	2-Aminoisobutyric acid	Aib	acide 2-amino-isobutyrique
bAib	3-Aminoisobuttersäure	bAib	3-Aminoisobutyric acid	bAib	acide 3-amino-isobutyrique
Apm	2-Aminopimelinsäure	Apm	2-Aminopimelic acid	Apm	acide 2-amino-pimélique
Dbu	2,4-Diaminobuttersäure	Dbu	2,4-Diaminobutyric acid	Dbu	acide 2,4-diaminobutyrique
Des	Desmosin	Des	Desmosine	Des	desmosine
Dpm	2,2'-Diaminopimelinsäure	Dpm	2,2'-Diaminopimelic acid	Dpm	acide 2,2'-diaminopimélique
Dpr	2,3-Diaminopropionsäure	Dpr	2,3-Diaminopropionic acid	Dpr	acide 2,3-diaminopropionique
EtGly	N-Ethylglycin	EtGly	N-Ethylglycine	EtGly	N-éthylglycine
EtAsn	N-Ethylasparagin	EtAsn	N-Ethylasparagine	EtAsn	N-éthylasparagine
Hyl	Hydroxylysin	Hyl	Hydroxylysine	Hyl	hydroxylysine
aHyl	allo-Hydroxylysin	aHyl	allo-Hydroxylysine	aHyl	allo-hydroxylysine
3Hyp	3-Hydroxyprolin	3Hyp	3-Hydroxyproline	3Hyp	3-hydroxyproline

4Hyp	4-Hydroxyprolin	4Hyp	4-Hydroxyproline	4Hyp	4-hydroxyproline
Ide	Isodesmosin	Ide	Isodesmosine	Ide	Isodesmosine
alle	allo-Isoleucin	alle	allo-Isoleucine	alle	allo-isoleucine
MeGly	N-Methylglycin, Sarkosin	MeGly	N-Methylglycine, sarcosine	MeGly	N-méthylglycine, sarcosine
Melle	N-Methylisoleucin	Melle	N-Methylisoleucine	Melle	N-méthylisoleucine
MeLys	6-N-Methyllysin	MeLys	6-N-Methyllysine	MeLys	6-N-méthyllysine
MeVal	N-Methylvalin	MeVal	N-Methylvaline	MeVal	N-méthylvaline
Nva	Norvalin	Nva	Norvaline	Nva	norvaline
Nle	Norleucin	Nle	Norleucine	Nle	norleucine
Orn	Ornithin	Orn	Ornithine	Orn	ornithine
SONSTIGE		OTHER		AUTRES	

2.8 Die den Codonen der codierenden Teile einer Nucleotidsequenz entsprechenden Aminosäuren sind unmittelbar unter die jeweiligen Codonen zu schreiben.

Wird ein Codon durch ein Intron aufgespalten, so ist das Aminosäuresymbol unter den Teil des Codons zu schreiben, der zwei Nucleotide enthält.

2.9 Die Basen einer Nucleotidsequenz (einschließlich Introns) sind jeweils in 10er-Gruppen aufzuführen; dies gilt nicht für die codierenden Teile der Sequenz.

Bleiben am Ende nichtcodierender Teile einer Sequenz weniger als 10 Basen als Rest, so sind diese zu einer Gruppe zusammenzufassen und durch einen Leerraum von angrenzenden Gruppen zu trennen.

2.10 Die Basen der codierenden Teile der Sequenz sind als Triplets (Codonen) aufzuführen.

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

2.12 Bei Nucleotidsequenzen sind höchstens 16 Codonen oder 60 Basen pro Zeile - mit einem Leerraum zwischen den einzelnen Codonen oder Gruppen von jeweils 10 Basen - aufzuführen.

2.13 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

2.8 Amino acids corresponding to the codons in the coding parts of a nucleotide sequence should be typed immediately under the corresponding codons.

Where a codon is split by an intron, the amino acid symbol should be typed below the portion of the codon containing two nucleotides.

2.9 The bases of a nucleotide sequence (including introns) should 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 noncoding parts of a sequence, shall be grouped together and separated from adjacent groups by a space.

2.10 The bases of the coding parts of a sequence should be listed as triplets (codons).

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

2.12 A nucleotide sequence should be listed with a maximum of 16 codons or 60 bases per line, with a space between each codon or group of 10 bases.

2.13 The enumeration of the nucleotide bases should start at the first base of the sequence with number 1. It should be continuous through the whole sequence in the direction 5' to 3'. It should 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 ap-

2.8 Les acides aminés correspondant aux codons dans les régions codantes d'une séquence nucléotidique devraient figurer immédiatement sous les codons correspondants.

Lorsqu'un codon est coupé par un intron, le symbole d'acide aminé devrait figurer sous la partie du codon contenant deux nucléotides.

2.9 Les bases d'une séquence nucléotidique (y compris les introns) devraient figurer sur la liste par groupes de dix bases, sauf celles situées dans les régions codantes de la séquence.

Les bases (moins de dix) qui restent à l'extrémité des régions non codantes d'une séquence devraient être regroupées et séparées des groupes voisins par un espace.

2.10 Les bases des régions codantes d'une séquence devraient figurer sur la liste sous forme de triplets (codons).

2.11 Une séquence de protéines ou de peptides devrait comporter seize acides aminés par ligne au maximum, avec un espace entre chaque acide aminé.

2.12 Une séquence nucléotidique devrait comporter seize codons ou soixante bases par ligne au maximum, avec un espace entre chaque codon ou groupe de dix bases.

2.13 L'énumération des bases nucléotidiques devrait commencer par la première base de la séquence, qui portera le numéro 1. Elle devrait être continue dans toute la séquence dans le sens 5' - 3'. Elle devrait figurer dans la marge de droite dans 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 nucléotidiques

Nucleotidsequenzen mit ringförmiger Konfiguration, wobei die Bestimmung der ersten Base der Nucleotidsequenz dem Anmelder überlassen bleibt.

plicable to nucleotide sequences that are circular in configuration, with the exception that the designation of the first base of the nucleotide sequence may be made at the option of the applicant.

s'applique également aux séquences nucléotidiques de configuration circulaire, à cette différence près que la désignation de la première base de la séquence nucléotidique peut être laissée au choix du demandeur.

2.14 Die Zählung der Aminosäuren beginnt bei der ersten Aminosäure des Proteins mit 1. Die dem reifen Protein vorausgehenden Aminosäuren wie beispielsweise Prä-Sequenzen, Pro-Sequenzen, Prä-Pro-Sequenzen und Signalsequenzen sind, soweit vorhanden, mit negativen Vorzeichen zu nummerieren, wobei die rückläufige Zählung mit der Aminosäure vor Nummer 1 beginnt.

2.14 The enumeration of amino acids should start at the first amino acid of the protein, with number 1. The amino acids preceding the mature protein, for example pre-sequences, pro-sequences, pre-pro-sequences and signal sequences, when present, should have negative numbers, counting backwards starting with the amino acid next to number 1.

2.14 L'énumération des acides aminés devrait commencer par le premier acide aminé de la protéine, qui portera le numéro 1. Les acides aminés précédant la protéine mature, soit par exemple les préséquences, les proséquences et les pré-proséquences ainsi que les séquences signal lorsqu'elles existent, devraient porter des nombres négatifs numérotés à rebours, en commençant par l'acide aminé voisin de l'acide portant le numéro 1.

Ansonsten beginnt die Zählung der Aminosäuren bei der ersten Aminosäure am Aminoterminus mit 1.

Otherwise, the enumeration of amino acids should start at the first amino acid at the amino terminal as number 1.

Sinon l'énumération des acides aminés devrait commencer par le premier acide aminé au niveau du terminal amino et porter le numéro 1.

Die Nummern sind im Abstand von jeweils 5 Aminosäuren unter der Sequenz anzugeben.

It should be marked under the sequence every 5 amino acids.

Le nombre devra figurer dans la séquence tous les cinq acides.

Die vorstehend beschriebene Zählweise für Aminosäuresequenzen gilt auch für Aminosäuresequenzen mit ringförmiger Konfiguration.

The enumeration method for amino acid sequences set forth above remains applicable for amino acid sequences that are circular in configuration.

La méthode présentée ci-dessus pour énumérer les séquences d'acides aminés s'applique également aux séquences d'acides aminés de configuration circulaire.

2.15 Teilsequenzen - die aus einem oder mehreren nichtbenachbarten Segmenten einer größeren Sequenz oder aus Segmenten verschiedener Sequenzen bestehen - sind als gesonderte Sequenzen mit eigener Sequenzkennzahl zu nummerieren. Sequenzen mit einer oder mehreren Lücken sind als mehrere gesonderte Sequenzen mit eigenen Sequenzkennzahlen zu nummerieren, wobei die Zahl der gesonderten Sequenzen der Zahl der jeweils zusammenhängenden Sequenzdatenreihen entspricht.

2.15 A partial sequence - made up of one or more non-contiguous segments of a larger sequence or of segments from different sequences - should be numbered as a separate sequence, with a separate sequence identifier. A sequence with a gap or gaps should 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.

2.15 Une séquence partielle, composée d'un segment ou de plusieurs segments non contigus d'une séquence plus grande ou de segments provenant de différentes séquences, devrait être numérotée comme une séquence distincte et porter un numéro d'identification de séquence distinct. Une séquence comportant un ou des espaces devrait être numérotée comme une série de séquences distinctes et porter des numéros d'identification distincts, le nombre de séquences distinctes étant égal au nombre de chaînes continues.

### 3. Sonstige verfügbare Angaben

Zusätzlich wird empfohlen, -soweit zutreffend und dem Anmelder bekannt - folgende Angaben einzureichen und unmittelbar vor der Sequenz selbst in das Sequenzprotokoll aufzunehmen. In diesem Zusammenhang sei betont, daß ein starkes Interesse am Zugang zu diesen zusätzlichen Daten besteht. Nähere Erläuterungen zu einer Sequenz wären eine sehr nützliche Informationsquelle in einer Patentsequenzdatenbank.

### 3. Other available Information

It is recommended that the following items of information be submitted, if applicable and if available to the applicant, and added to the Sequence Listing immediately prior to the sequence itself. It should be emphasized that there is a strongly perceived need for access to these additional data. Detailed annotations of the sequence would be a very useful source of information in a patent sequence database.

### 3. Autres renseignements

Il est recommandé au demandeur de fournir les informations suivantes, si elles sont pertinentes et s'il en dispose, et de les ajouter sur la liste des séquences juste avant la séquence elle-même. Il convient de souligner que l'accès à ces données supplémentaires est un besoin fortement ressenti. Des annotations détaillées afférentes à la séquence constitueraient une source d'information très utile dans une base de données relatives aux séquences contenues dans les brevets.

**3.1 Strangform:**

Nucleotidsequenzen sollten durch eine der folgenden Angaben näher spezifiziert werden:

- Einzelstrang
- Doppelstrang
- beides
- dem Anmelder nicht bekannt

**3.2 Topologie der Sequenz:**

Hier sollte die Konfiguration der Sequenz angegeben werden:

- linear
- ringförmig
- beides
- Topologie dem Anmelder nicht bekannt

**3.3 Art des sequenzierten Moleküls:**

Das Sequenzprotokoll sollte mindestens eine der folgenden Angaben enthalten:

- Genom-DNA
- Genom-RNA
- m-RNA
- t-RNA
- r-RNA
- sn-RNA
- sc-RNA
- prä-RNA
- c-DNA zu Genom-RNA
- c-DNA zu m-RNA
- c-DNA zu t-RNA
- c-DNA zu r-RNA
- c-DNA zu sn-RNA
- c-DNA zu sc-RNA
- sonstige Nucleinsäure (angeben)
- Peptid
- Protein

**3.4 Hypothetische Sequenzen:**

Hier ist anzugeben, ob es sich um eine hypothetische Sequenz handelt oder nicht.

**3.5 Antisense:**

Hier ist anzugeben, ob es sich um eine Antisense-Sequenz handelt oder nicht.

**3.6 Art des Fragments - nur bei Proteinen oder Peptiden:**

In das Sequenzprotokoll ist mindestens eine der folgenden Angaben aufzunehmen:

- N-Terminus
- C-Terminus
- inneres Fragment

**3.1 The strandedness:**

If the sequence is a nucleotide sequence, indicate whether it is:

- single-stranded
- double-stranded
- both
- unknown to applicant

**3.2 The topology of the sequence:**

Indicate whether the sequence is:

- linear
- circular
- both
- topology unknown to applicant

**3.3 The type of molecule sequenced:**

At least one of the following should be included in the Sequence Listing,

- genomic DNA
- genomic RNA
- mRNA
- tRNA
- rRNA
- snRNA
- scRNA
- preRNA
- cDNA to genomic RNA
- cDNA to mRNA
- cDNA to tRNA
- cDNA to rRNA
- cDNA to snRNA
- cDNA to scRNA
- other nucleic acid (identify)
- peptide
- protein

**3.4 Hypothetical sequences:**

Indicate whether the sequence is hypothetical or not.

**3.5 Anti-sense:**

Indicate whether the sequence is anti-sense or not.

**3.6 The fragment type for proteins or peptides only:**

At least one of the following should be included in the Sequence Listing:

- N-terminal
- C-terminal
- internal fragment

**3.1 Nombre de brins :**

S'il s'agit d'une séquence nucléotidique, il convient d'indiquer si elle est :

- à simple brin
- à double brin
- les deux
- non connue du demandeur.

**3.2 Configuration de la séquence :**

Il convient d'indiquer si la séquence est :

- linéaire
- circulaire
- les deux
- de configuration non connue du demandeur.

**3.3 Type de molécule séquencée :**

L'un au moins des types suivants devra figurer sur la liste des séquences :

- ADN génomique
- ARN génomique
- ARNm
- ARNt
- ARNr
- ARNsn
- ARNsc
- pré ARN
- ADNc pour ARN génomique
- ADNc pour ARNm
- ADNc pour ARNt
- ADNc pour ARNr
- ADNc pour ARNsn
- ADNc pour ARNsc
- autre acide nucléique (à préciser)
- peptide
- protéine

**3.4 Séquences hypothétiques**

Préciser si la séquence est hypothétique ou non.

**3.5 Anti-sens**

Préciser si la séquence est anti-sens ou non.

**3.6 Type de fragment, uniquement pour les protéines ou les peptides :**

L'un au moins des types de fragment devrait figurer sur la liste des séquences:

- N-terminal
- C-terminal
- fragment interne

3.7 Ursprüngliche Herkunft des sequenzierten Moleküls:	3.7 The original source of the molecule sequenced:	3.7 Origine de la molécule
Weitere Angaben können zu folgenden Punkten gemacht werden:	More information may be included on:	Des précisions peuvent être apportées en ce qui concerne :
<ul style="list-style-type: none"> <li>- Name des Organismus (Art)</li> <li>- Stamm</li> <li>- Name/Nummer des Individuums oder Isolats</li> <li>- Entwicklungsstadium [ ] Keimlinie [ ] rearranged</li> <li>- Hapbtyp</li> <li>- Gewebetyp</li> <li>- Zelltyp</li> <li>- Zelllinie</li> <li>- Organelle</li> </ul>	<ul style="list-style-type: none"> <li>- organism (species) name</li> <li>- strain</li> <li>- name/number of individual or isolate</li> <li>- development stage [ ] germ line [ ] rearranged</li> <li>- haplotype</li> <li>- tissue type</li> <li>- cell type</li> <li>- cell line</li> <li>- organelle</li> </ul>	<ul style="list-style-type: none"> <li>- nom de l'organisme (espèce)</li> <li>- souche</li> <li>- nom/numéro de l'organisme individuel ou isolé</li> <li>- stade de développement [ ] lignée germinale [ ] réagencé</li> <li>- haplotype</li> <li>- type de tissu</li> <li>- type de cellule</li> <li>- lignée cellulaire</li> <li>- organelle</li> </ul>
3.8 Unmittelbare <b>experimentelle Herkunft</b> der Sequenz:	3.8 The immediate <b>experimental source</b> of the sequence:	3.8 <b>Source expérimentale</b> immédiate de la séquence :
<ul style="list-style-type: none"> <li>- Bibliothek (Typ, Name)</li> <li>- Clon(e)</li> </ul>	<ul style="list-style-type: none"> <li>- library (type, name)</li> <li>- clone(s)</li> </ul>	<ul style="list-style-type: none"> <li>- bibliothèque (type, nom)</li> <li>- clone(s)</li> </ul>
3.9 <b>Position</b> der Sequenz im Genom:	3.9 <b>The position</b> of the sequence in the genome:	3.9 <b>Position</b> de la séquence dans le génome :
<ul style="list-style-type: none"> <li>- Name/Nummer des Chromosoms (oder Segments)</li> <li>- Kartenposition</li> <li>- Einheiten: [ ] Genom % oder [ ] Nucleotidnummer oder [ ] sonstige (bitte angeben)</li> </ul>	<ul style="list-style-type: none"> <li>- chromosome (or segment) name/number</li> <li>- map position</li> <li>- units: [ ] genome % or [ ] nucleotide number or [ ] other (specify)</li> </ul>	<ul style="list-style-type: none"> <li>- nom/numéro du chromosome (ou segment)</li> <li>- position sur la carte :</li> <li>- unités : [ ] % génome ou [ ] numéro du nucléotide ou [ ] autre (à préciser)</li> </ul>
3.10 <b>Merkmale</b> der Sequenz:	3.10 <b>Features</b> of the sequence:	3.10 <b>Caractéristiques</b> de la séquence
- <b>Lage der Merkmale:</b>	- <b>location of the features:</b>	- <b>emplacement des caractéristiques</b>
von (Nummer der ersten Base/Aminosäure des Merkmals)	from (number of first base/amino acid in the feature)	de (numéro de la première base/du premier acide aminé dans la caractéristique)
bis (Nummer der letzten Base/Aminosäure des Merkmals) Basenpaare (Ziffern verweisen auf die Positionen der Basenpaare in einer Nucleotidsequenz) Aminosäuren (Ziffern verweisen auf die Positionen der Aminosäurereste in einer Peptidsequenz)	to (number of last base/amino acid in the feature) base pairs (numbers refer to positions of base pairs in a nucleotide sequence) amino acids (numbers refer to positions of amino acid residues in a peptide sequence)	à (numéro de la dernière base/du dernier acide aminé dans la caractéristique) paires de bases (les numéros renvoient aux positions des paires de bases dans une séquence nucléotidique) acides aminés (les numéros renvoient aux positions des résidus d'acide aminé dans une séquence de peptides).
mit Angaben darüber, ob sich das Merkmal auf dem zum Strang des Sequenzprotokolls komplementären Strang befindet	including whether feature is located on the complementary strand to that filed in the Sequence Listing	Déterminer également si la caractéristique se situe sur le brin complémentaire de celui figurant dans la liste des séquences.
- <b>Art der Ermittlung des Merkmals:</b>	- <b>identification method by which the feature was identified:</b>	- <b>méthode d'identification de la caractéristique</b>
E = experimentell S = durch Ähnlichkeit mit einer bekannten Sequenz oder einer feststehenden Consensus-Sequenz P = durch Ähnlichkeit mit einem anderen Schema	E = experimentally S = by similarity with a known sequence or to an established consensus sequence P = by similarity to some other pattern	E = expérimentale S = par similitude avec une séquence connue ou avec une séquence consensus connue P = par similitude avec d'autres modèles

**- Merkmalsschlüssel:**

Entsprechend dem in den bestehenden Datenbanken für Nichtpatentliteratur eingeführten Merkmalsvokabular gemäß der DDBJ/EMBL/GenBank Feature Tables Definition für Merkmalstabellen können signifikante Merkmale folgendes umfassen:

Allele  
Attenuator  
CAAT-Signal  
zellulär  
codierende Sequenz  
Konflikt  
D-loop Enhancer  
Exon  
GC-Signal  
i-DNA  
Insertionssequenz  
Intron  
LTR long terminal repeat  
reifes Peptid  
modifizierte Base  
m-RNA  
Mutation  
polyA-Signal  
Vorläufer-RNA  
Primärtranskript  
primerbindend  
Promotor  
proteinbindend  
Provirus  
RBS Ribosomenbindungsstelle  
Repetitionseinheit  
repetitiver Abschnitt  
Replikationsursprung  
r-RNA  
Satellit  
sc-RNA  
Signalpeptid  
sn-RNA  
stem loop  
TATA-Signal  
Terminator  
Transit-Peptid  
Transposon  
t-RNA  
unsicher  
Variation  
Virion  
3'Clip  
3'UTR  
5'Clip  
5'UTR  
-10-Signal  
-35-Signal  
oder sonstige

**- sonstige bedeutsame Angaben:**

- assoziierter Phänotyp(en)
- biologische/enzymatische Wirkung
- biologische/enzymatische Wirkung des Produkts

**- features keys:**

Following the controlled vocabulary of the features field in the existing non-patent literature databases according to the DDBJ/EMBL/GenBank Feature Tables Definition, significant features might include:

allele  
attenuator  
CAAT signal  
cellular  
coding sequence  
conflict  
D-loop enhancer  
exon  
GC signal  
iDNA  
insertion sequence  
intron  
LTR long terminal repeat  
mature peptide  
modified base  
mRNA  
mutation  
polyA signal  
precursor RNA  
primary transcript  
primer binding  
promoter  
protein binding  
provirus  
RBS ribosome binding site  
repeating unit  
repeat region  
replication origin  
rRNA  
Satellite  
scRNA  
signal peptide  
snRNA  
stem loop  
TATA signal  
terminator  
transit peptide  
transposon  
tRNA  
unsure  
variation  
virion  
3'clip  
3'UTR  
5'clip  
5'UTR  
-10 signal  
-35 signal  
or others

**- other Information of significance:**

- associated phenotype(s)
- biological/enzymatic activity
- biological/enzymatic activity of its product

**- caractéristiques Importantes**

Conformément à la terminologie établie dans le domaine des caractéristiques dans les bases de données non-brevets existantes répondant à la DDBJ/EMBL/GenBank Feature Table Definition, les caractéristiques importantes peuvent inclure :

allèle  
atténuateur  
signal CAAT  
cellulaire  
séquence codante  
boucle D en opposition  
activateur  
exon  
signal GC  
ADNi  
séquence d'insertion  
intron  
longue région terminale répétée LTR  
peptide mature  
base modifiée  
ARNm  
mutation  
signal poly A  
précurseur ARN  
transcript primaire  
fixation de l'amorce  
promoteur  
fixation de la protéine  
provirus  
site de fixation du ribosome  
unité répétée  
région répétée  
origine de répliation  
ARNr  
satellite  
ARNsc  
peptide signal  
ARNsn  
structure en épingle à cheveux  
signal TATA  
termineur  
peptide de transit  
transposon  
ARNt  
indéterminé  
variation  
virion  
3'clip  
3'UTR  
5'clip  
5'UTR  
signal -10  
signal -35  
ou autres

**- autres renseignements Importants**

- phénotype(s) associé(s)
- activité biologique/enzymatique
- activité biologique/enzymatique de son produit



- allgemeine Funktionsklasse des Gens und/oder Genprodukts
  - Makromolekülbindung - Makromoleküle, an die sich das Genprodukt binden kann
  - subzelluläre Lokalisation - subzelluläre Lokalisation des Genprodukts
  - sonstige sachdienliche Angaben
- general functional class of the gene and/or gene product
  - binding macromolecules - macromolecules to which the gene product can bind
  - subcellular localisation - subcellular localisation of the gene product
  - any other relevant information
- classe de fonction générale du gène et/ou produit du gène
  - macromolécules de fixation - macromolécules auxquelles le produit du gène peut se fixer
  - localisation subcellulaire - localisation subcellulaire du produit du gène
  - toute autre information pertinente

## ANLAGE 2

## ANNEX 2

## ANNEXE 2

## Muster eines Sequenzprotokolls

## Specimen Sequence Listing

## Exemple de Liste des séquences

SEQ ID NO: 1  
ART DER SEQUENZ: Nucleotid mit  
entsprechendem Protein  
SEQUENZLÄNGE: 2654 Basenpaare

STRANGFORM: Einzelstrang  
TOPOLOGIE: linear  
ART DES MOLEKÜLS: Genom-DNA

URSPRÜNGLICHE HERKUNFT  
ORGANISMUS: Mensch  
UNMITTELBARE EXPERIMENTELLE  
HERKUNFT  
NAME DER ZELLINIE: humane Leuko-  
cyten von gesunden Erwachsenen

MERKMALE:  
von 339 bis 344 BP TATA-Signal

von 527 bis 814 BP Intron 1  
von 923 bis 1006 BP Intron 2  
von 1113 bis 1359 BP Intron 3  
von 824 bis 1009 BP Signalpeptid  
von 1010 bis 1772 BP reifes Peptid

EIGENSCHAFTEN: humanes Lympho-  
toxin-Gen

SEQ ID NO: 1  
SEQUENCE TYPE: Nucleotide with  
corresponding protein  
SEQUENCE LENGTH: 2654 base  
pairs

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA

ORIGINAL SOURCE  
ORGANISM: human  
IMMEDIATE EXPERIMENTAL  
SOURCE  
NAME OF CELL LINE: human leu-  
cocyte from healthy adults

FEATURES:  
from 339 to 344 bp TATA Signal

from 527 to 814 bp intron 1  
from 923 to 1006 bp intron 2  
from 1113 to 1359 bp intron 3  
from 824 to 1009 bp signal peptide  
from 1010 to 1772 bp mature  
peptide

PROPERTIES: human lymphotoxin  
gene

SEQ ID NO: 1  
TYPE DE SEQUENCE : Nucléotide et  
sa protéine correspondante  
LONGUEUR DE LA SEQUENCE :  
2654 paires de bases

NOMBRE DE BRINS : simple  
CONFIGURATION : linéaire  
TYPE DE MOLECULE : ADN  
génomique

ORIGINE  
ORGANISME : humain  
SOURCE EXPERIMENTALE IMMEDI-  
ATE  
NOM DE LA LIGNEE CELLULAIRE :  
leucocytes humains provenant  
d'adultes sains

CHARACTERISTIQUES :  
de 339 à 344 paires de bases signal  
TATA  
de 527 à 814 paires de bases intron 1  
de 923 à 1006 paires de bases intron 2  
de 1113 à 1359 paires de bases intron 3  
de 824 à 1009 paires de bases peptide  
signal de 1010 à 1772 paires de bases  
peptide mature

PROPRIETES : gène humain de la  
lymphotoxine

AAGGGTGCAG	AGATGTTATA	TATGATTGCT	CTTCAGGGAA	CCGGCCTCCA	GCTCACACCC	60
CAGCTGCTCA	ACCGCCTCCT	CTCTGAATTG	ACTGTCCCTT	CTTTGGAAct	CTAGGCCTGA	120
CCCCACTCCC	TGGCCCTCCC	AGCCCACGAT	TCCCCTGACC	CGACTCCCTT	TCCCAGAAct	180
CAGTCGCCTG	AACCCCCAGC	CTGTGGTTCT	CTCCTAGGCC	TCAGCCTTTC	CTGCCTTTGA	240
CTGAAACAGC	AGTATCTTCT	ACACGCTGGG	GCTTCCC GCG	GCCCAGCCCC	GACCTAGAAC	300
CCGCCCGCTG	CCTGCCACGC	TGCCACTGCC	GCTTCTCTA	TAAAGGGACC	TGAGCGTCCG	360
CGCGCAGGGG	CTCCACACAG	CAGGTGAGGC	TCTCCTGCCC	CATCTCCTTG	GGCTGCCCGT	420
GCTTCGTGCT	TTGGACTACC	GCCCCGAGTG	TCCTGCCCTC	TGCCTGGGCC	TCGGTCCCTC	480
CTGCACCTGC	TGCCTGGATC	CCCGGCCTGC	CTGGGCCTGG	GCCTTGGTGG	GTTTGGTTTT	540
GGTTTCCTTC	TCTGTCTCTG	ACTCTCCATC	TGTCAGTCTC	ATTGTCTCTG	TCACACATTC	600
TCTGTTTCTG	CCATGGTTCC	TCTCTGTTCC	CTTCCTGTCT	CTCTCTGTCT	CCCTCTGCTC	660
ACCTTGGGGT	TTCTCTGACT	GCATCTTGTC	CCCTTCTCTG	TCCGATCTCT	CTCTCGGGGG	720
TCGGGGGGTG	CTGTCTCCCA	GGGCGGGAGG	TCTGTCTTCC	GCCGCGTGCC	CCGCCCCGCT	780
CACTGTCTCT	CTCTCTCTCT	CTCTTTCTCT	GCAGTTTCTC	CCC ATG ACA	CCA CCT	835
				Met Thr Pro Pro		

GAA CGT CTC	TTC CTC CCA	AGG GTG TGT	GGC ACC ACC	CTA CAC CTC	CTC CTC	883
Glu Arg Leu	Phe Leu Pro	Arg Val Cys	Gly Thr Thr	Leu His Leu	Leu Leu	
-30	-25		-20		-15	

CTT CTG GGG	CTG CTG CTG	GTT CTG CTG	CCT GGG GCC	CAG GTGAGGCAGC		932
Leu Leu Gly	Leu Leu Leu	Val Leu Leu	Pro Gly Ala	Gln		
	-10		-5			

AGGAGAATGG	GGGCTGCTGG	GGTGGCTCAG	CCAAACCTTG	AGCCCTAGAG	CCCCCTCAA	992
------------	------------	------------	------------	------------	-----------	-----

CTCTGTTCTC	CTAG	GGG	CTC	CCT	GGT	GTT	GGC	CTC	ACA	CCT	TCA	GCT	GCC	1042		
		Gly	Leu	Pro	Gly	Val	Gly	Leu	Thr	Pro	Ser	Ala	Ala			
			1				5					10				
CAG	ACT	GCC	CGT	CAG	CAC	CCC	AAG	ATG	CAT	CTT	GCC	CAC	AGC	AAC	CTC	1090
Gln	Thr	Ala	Arg	Gln	His	Pro	Lys	Met	His	Leu	Ala	His	Ser	Asn	Leu	
			15					20					25			
AAA	CCT	GCT	GCT	CAC	CTC	ATT	GTAAACATCC	ACCTGACCTC	CCAGACATGT							1142
Lys	Pro	Ala	Ala	His	Leu	Ile										
			30													
CCCCACCAGC	TCTCCTCCTA	CCCCTGCCTC	AGGAACCCAA	GCATCCACCC	CTCTCCCCCA									1202		
ACTTCCCCCA	CGCTAAAAAA	AACAGAGGGA	GCCCCTCCT	ATGCCTCCCC	CTGCCATCCC									1262		
CCAGGAACTC	AGTTGTTTCCAG	TGCCCACTTC	CTCAGGGATT	GAGACCTCTG	ATCCAGACCC									1322		
CTGATCTCCC	ACCCCCATCC	CCTATGGCTC	TTCCTAG	GA	GAC	CCC	AGC	AAG	CAG					1376		
								Gly	Asp	Pro	Ser	Lys	Gln			
								35					40			
AAC	TCA	CTG	CTC	TGG	AGA	GCA	AAC	ACG	GAC	CGT	GCC	TTC	CTC	CAG	GAT	1424
Asn	Ser	Leu	Leu	Trp	Arg	Ala	Asn	Thr	Asp	Arg	Ala	Phe	Leu	Gln	Asp	
			45						50					55		
GGT	TTC	TCC	TTG	AGC	AAC	AAT	TCT	CTC	CTG	GTC	CCC	ACC	AGT	GGC	ATC	1472
Gly	Phe	Ser	Leu	Ser	Asn	Asn	Ser	Leu	Leu	Val	Pro	Thr	Ser	Gly	Ile	
			60					65						70		
TAC	TTC	GTC	TAC	TCC	CAG	GTG	GTC	TTC	TCT	GGG	AAA	GCC	TAC	TCT	CCC	1520
Tyr	Phe	Val	Tyr	Ser	Gln	Val	Val	Phe	Ser	Gly	Lys	Ala	Tyr	Ser	Pro	
		75				80						85				
AAG	GCC	ACC	TCC	TCC	CCA	CTC	TAC	CTG	GCC	CAT	GAG	GTC	CAG	CTC	TTC	1568
Lys	Ala	Thr	Ser	Ser	Pro	Leu	Tyr	Leu	Ala	His	Glu	Val	Gln	Leu	Phe	
	90					95					100					
TCC	TCC	CAG	TAC	CCC	TTC	CAT	GTG	CCT	CTC	CTC	AGC	TCC	CAG	AAG	ATG	1616
Ser	Ser	Gln	Tyr	Pro	Phe	His	Val	Pro	Leu	Leu	Ser	Ser	Gln	Lys	Met	
105					110					115					120	
GTG	TAT	CCA	GGG	CTG	CAG	GAA	CCC	TGG	CTG	CAC	TCG	ATG	TAC	CAC	GGG	1664
Val	Tyr	Pro	Gly	Leu	Gln	Glu	Pro	Trp	Leu	His	Ser	Met	Tyr	His	Gly	
			125					130					135			
GCT	GCG	TTC	CAG	CTC	ACC	CAG	GGA	GAC	CAG	CTA	TCC	ACC	CAC	ACA	GAT	1712
Ala	Ala	Phe	Gln	Leu	Thr	Gln	Gly	Asp	Gln	Leu	Ser	Thr	His	Thr	Asp	
			140				145					150				
GGC	ATC	CCC	CAC	CTA	GTC	CTC	AGC	CCT	AGT	ACT	GTC	TTC	TTT	GGA	GCC	1760
Gly	Ile	Pro	His	Leu	Val	Leu	Ser	Pro	Ser	Thr	Val	Phe	Phe	Gly	Ala	
		155					160							165		

