**NORME ST.26 – ANNEXE vI**

Guide d’application

*Version 1,*~~1~~*2*

*Projet final*

~~Adopté par le Comité des normes de l’OMPI (CWS)  
à sa cinquième session du 2 juin 2017~~Proposition présentée par l’équipe d’experts SEQL pour examen et approbation par le CWS à sa sixième session

## **Introduction**

La présente norme a notamment pour but de “permettre aux déposants d’établir, dans le cadre d’une demande de brevet, un listage des séquences unique qui soit acceptable pour les procédures internationales et nationales ou régionales”. Le présent document d’orientation vise à faire en sorte que tous les déposants et offices de la propriété intellectuelle comprennent et acceptent les exigences en matière d’intégration et de représentation des divulgations de séquences, de sorte que ce but puisse être atteint.

Le présent document comprend la présente introduction, un index des exemples, des exemples de divulgation de séquences et un appendice contenant un listage des séquences en XML établi à partir des séquences tirées des exemples. La présente introduction explique certains concepts et termes utilisés dans le reste du document. Les exemples illustrent les exigences des différents paragraphes de la norme et il est renvoyé pour chaque exemple au numéro de paragraphe qui correspond le mieux à l’objet traité. Par ailleurs, certains exemples illustrent d’autres paragraphes, et des renvois appropriés sont indiqués à la fin de chaque exemple. L’index présente les numéros de page des exemples, avec indications des éventuels renvois. Chaque séquence d’un exemple qui doit ou peut être intégrée dans un listage des séquences s’est vu attribuer un numéro d’identification de séquence (SEQ ID NO) et apparaît au format XML dans l’[Appendice](#Appendix) au présent document.

Pour chaque exemple, toute information donnée à titre d’explication d’une séquence doit être considérée comme l’intégralité de la divulgation concernant cette séquence. Les réponses données ne tiennent compte que des informations présentées de manière explicite dans l’exemple.

Les orientations fournies dans le présent document portent sur l’établissement d’un listage des séquences à fournir à la **date du dépôt** d’une demande de brevet. Quant à l’établissement d’un listage des séquences à fournir **après la date du dépôt** d’une demande de brevet, il convient de se demander si les informations fournies pourraient être considérées par un office de la propriété intellectuelle comme ajoutant des éléments à la divulgation originale. Il s’ensuit que les orientations fournies dans le présent document pourront ne pas être applicables à un listage des séquences présenté après la date du dépôt d’une demande de brevet.

### Établissement d’un listage des séquences

Pour établir un listage des séquences, il importe de se poser les questions suivantes :

1. Le paragraphe 7 de la norme ST.26 prescrit‑il l’intégration d’une séquence divulguée donnée?

2. Si l’intégration d’une séquence divulguée donnée n’est pas prescrite, l’intégration de cette séquence est‑elle autorisée par la norme ST.26?

3. Si l’intégration d’une séquence divulguée donnée est prescrite ou autorisée par la norme ST.26, comment cette séquence doit‑elle être représentée dans le listage des séquences?

S’agissant de la première question, le paragraphe 7 (avec certaines restrictions) de la norme ST.26 prescrit l’intégration d’une séquence qui est divulguée dans une demande de brevet par **l’énumération de ses résidus**, lorsque la séquence contient au moins 10 nucléotides **définis de manière spécifique** ou au moins quatre acides aminés **définis de manière spécifique**.

S’agissant de la deuxième question, le paragraphe 8 de la norme ST.26 interdit l’intégration d’une séquence comportant moins de 10 nucléotides **définis de manière spécifique** ou moins de quatre acides aminés **définis de manière spécifique**.

Pour répondre à ces deux questions, il importe de bien comprendre l’“énumération de ses résidus” et “définis de manière spécifique”.

S’agissant de la troisième question, le présent document contient les divulgations de séquences qui illustrent différents scénarios, ainsi qu’un examen exhaustif du moyen de représentation préféré de chaque séquence ou, lorsqu’une séquence présente de multiples variantes, de la “**séquence la plus englobante**”, conformément à la présente norme. Vu l’impossibilité de prendre en compte tous les scénarios de séquences inhabituelles possibles, le présent document d’orientation s’efforce de présenter le raisonnement qui sous‑tend l’approche de chaque exemple et la manière dont les dispositions de la norme ST.26 sont appliquées, de sorte que le même raisonnement peut être appliqué aux autres scénarios de séquences non illustrés.

#### Énumération de ses résidus

Le paragraphe 3.c) de la norme ST.26 définit l’expression **“énumération de ses résidus”** comme désignant la divulgation d’une séquence dans une demande de brevet sous forme de listage, dans un ordre donné, de chacun des résidus de la séquence, étant entendu que i) le résidu est représenté par un nom, une abréviation, un symbole ou une structure; ou ii) les résidus multiples sont représentés par une formule topologique. Une séquence devrait être divulguée dans une demande de brevet par l’“énumération de ses résidus” à l’aide de symboles conventionnels, qui sont les symboles des nucléotides indiqués dans le tableau 1 de la section 1 de l’annexe 1 à la norme ST.26 (c’est‑à‑dire les symboles en lettres minuscules ou leurs équivalents en lettres majuscules) et les symboles des acides aminés indiqués dans le tableau 3 de la section 3 de l’annexe 1 à la norme ST.26 (c’est‑à‑dire les symboles en lettres majuscules ou leurs équivalents en lettres minuscules)[[1]](#footnote-2). Les symboles autres que ceux indiqués dans ces tableaux sont “**non conventionnels**”.

Selon une méthode non préférée, une séquence est parfois divulguée par l’“énumération de ses résidus” à l’aide d’**abréviations conventionnelles** oude **noms complets** (par opposition à des symboles conventionnels) conformément à ce qui est indiqué dans les tableaux A et B plus loin, de symboles ou d’abréviations conventionnels utilisés d’une manière non conventionnelle, de symboles ou d’abréviations non conventionnels, de formules ou structures chimiques ou de formules topologiques. Il faudrait s’efforcer de divulguer les séquences selon la méthode préférée; toutefois, lorsque les séquences sont divulguées selon une méthode non préférée, il peut être nécessaire de consulter l’explication de la séquence donnée dans la divulgation afin de déterminer la signification du symbole ou de l’abréviation non préféré.

Lorsqu’un symbole ou une abréviation conventionnel est utilisé, on n’en doit pas moins consulter l’explication de la séquence donnée dans la divulgation afin de confirmer que le symbole est utilisé d’une manière conventionnelle. Autrement, si le symbole est utilisé d’une manière non conventionnelle, l’explication est nécessaire pour déterminer si le paragraphe 7 de la norme ST.26 prescrit l’intégration dans le listage des séquences ou si le paragraphe 8 interdit cette intégration.

Lorsqu’un symbole ou une abréviation non conventionnel est divulgué comme équivalant à un symbole ou une abréviation conventionnel (par exemple, “Z1” désigne “A”) ou à une séquence particulière de symboles conventionnels (par exemple, “Z1” désigne “agga”), la séquence est interprétée comme si elle était divulguée à l’aide du ou des symboles ou abréviations conventionnels équivalents, afin de déterminer si le paragraphe 7 de la norme ST.26 prescrit l’intégration dans le listage des séquences ou si le paragraphe 8 interdit cette intégration. Lorsqu’un symbole non conventionnel de nucléotide est utilisé comme symbole ambigu (par exemple, X1 = inosine ou pseudouridine), mais sans être l’équivalent de l’un des symboles ambigus conventionnels indiqués dans le tableau 1 de la section 1 (c’est‑à‑dire “m”, “r”, “w”, “s”, “y”, “k”, “v”, “h”, “d”, “b”, ou “n”), le résidu est interprété en tant que résidu “n” afin de déterminer si le paragraphe 7 de la norme ST.26 prescrit l’intégration dans le listage des séquences ou si le paragraphe 8 de la norme ST.26 interdit cette intégration. De même, lorsqu’un symbole non conventionnel d’acide aminé est utilisé comme symbole ambigu (par exemple “Z1” désigne “A”, “G”, “S” ou “T”), mais sans être l’équivalent de l’un des symboles ambigus conventionnels indiqués dans le tableau 3 de la section 3 (c’est‑à‑dire B, Z, J ou X), le résidu est interprété en tant que résidu “X” afin de déterminer si le paragraphe 7 de la norme ST.26 prescrit l’intégration de la séquence dans le listage des séquences ou si le paragraphe 8 de la norme ST.26 interdit cette intégration.

#### Définis de manière spécifique

Le paragraphe 3.k) de la norme ST.26 définit l’expression “**définis de manière spécifique**” comme désignant tout nucléotide différent de ceux qui sont représentés par le symbole “n” et tout acide aminé différent de ceux qui sont représentés par le symbole “X” dans l’annexe I, étant entendu que “n” et “X” sont utilisés d’une manière conventionnelle telle que décrite dans le tableau 1 de la section 1 (c’est‑à‑dire “a ou c ou g ou t/u; ‘unknown’ ou ‘other’”) et dans le tableau 3 de la section 3 (c’est‑à‑dire A ou R ou N ou D ou C ou Q ou E ou G ou H ou I ou L ou K ou M ou F ou P ou O ou S ou U ou T ou W ou Y ou V, ‘unknown’ ou ‘other’”), respectivement. Pour déterminer si un nucléotide ou un acide aminé est “défini de manière spécifique”, il sera tenu compte des indications susmentionnées concernant les symboles conventionnels ou les symboles ou abréviations non conventionnels et leur utilisation d’une manière conventionnelle ou non conventionnelle.

#### *Séquence la plus englobante*

Lorsqu’une séquence qui répond aux exigences du paragraphe 7 n’est divulguée par l’énumération de ses résidus qu’une seule fois dans une demande, mais est décrite d’une manière différente dans de multiples modes de réalisation – par exemple, dans un mode de réalisation, “X” pourrait être, à un ou plusieurs emplacements, tout acide aminé, mais, dans des modes de réalisation ultérieurs, “X” pourrait ne représenter qu’un nombre limité d’acides aminés –, la norme ST.26 prescrit l’intégration dans un listage des séquences uniquement de la séquence qui a été énumérée par ses résidus. Conformément aux paragraphes 15 et 27, lorsqu’une telle séquence contient des symboles ambigus multiples “n” ou “X”, “n” ou “X” est considéré comme représentant tout nucléotide ou acide aminé, respectivement, en l’absence d’annotation supplémentaire. En conséquence, la seule séquence à intégrer est la séquence la plus englobante divulguée. La **séquence la plus englobante** est la séquence unique dont les résidus de variante sont représentés par les symboles ambigus les plus restrictifs qui intègrent les modes de réalisation les plus divulgués. Toutefois, l’intégration de séquences particulières supplémentaires est fortement conseillée lorsqu’elle est possible, par exemple lorsque ces séquences représentent des modes de réalisation supplémentaires qui constituent une partie essentielle de l’invention. L’intégration des séquences supplémentaires permet d’effectuer une recherche plus approfondie et porte à la connaissance du public l’objet pour lequel la délivrance d’un brevet est demandée.

### Usage du symbole ambigu

#### Bon usage du symbole ambigu “n” dans un listage des séquences

Le symbole “n”

* 1. ne doit être employé que pour représenter un seul nucléotide;
  2. sera considéré comme l’équivalent de l’un des symboles “a”, “c”, “g” ou “t/u”, sauf s’il est accompagné d’une description supplémentaire;
  3. devrait être employé pour représenter l’un quelconque des nucléotides ci‑après s’il est accompagné d’une description supplémentaire :
     1. un nucléotide modifié, par exemple un nucléotide naturel, synthétique ou n’existant pas à l’état naturel, qui ne peut être représenté à l’aide d’un autre symbole indiqué dans l’annexe I (voir section 1, tableau 1);
     2. un nucléotide “unknown”, c’est‑à‑dire non déterminé, non divulgué ou incertain;
     3. un site abasique; ou
  4. peut être employé pour représenter une variante de séquence, c’est‑à‑dire des alternatives, des suppressions, des adjonctions ou des remplacements, lorsque “n” est le symbole ambigu le plus restrictif.

#### *Bon usage du symbole ambigu “X” dans un listage des séquences*

Le symbole “X”

1. ne peut être employé que pour représenter un acide aminé.
2. ne sera pas considéré comme équivalent à l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire
3. devrait être employé pour représenter l’un quelconque des acides aminés ci‑après s’il est accompagné d’une description supplémentaire :
   * 1. un acide aminé modifié, par exemple un acide aminé naturel, synthétique ou n’existant pas à l’état naturel, qui ne peut être représenté à l’aide d’un autre symbole indiqué dans l’annexe I (voir section 3, tableau 3);
     2. un acide aminé “unknown”, c’est‑à‑dire non déterminé, non divulgué ou incertain; ou
4. peut être employé pour représenter une variante de séquence, c’est‑à‑dire des alternatives, des suppressions, des adjonctions ou des remplacements, lorsque “X” est le symbole ambigu le plus restrictif.

|  |  |  |
| --- | --- | --- |
| **Symbole** | **Abréviation** | **Nom du nucléotide** |
| A |  | Adénine |
| C |  | Cytosine |
| G |  | Guanine |
| T |  | Thymine dans l’ADN  Uracile dans l’ARN (t/u) |
| M | a ou c |
| R | a ou g |
| w | a ou t/u |
| s | c ou g |
| y | c ou t/u |
| k | g ou t/u |
| V | a ou c ou g; et non t/u |
| h | a ou c ou t/u; et non g |
| d | a ou g ou t/u; et non c |
| b | c ou g ou t/u; et non a |
| n | a, c, g, t/u, “unknown” ou “other” |

### Tableau A – Symboles, abréviations et noms conventionnels des nucléotides

|  |  |  |
| --- | --- | --- |
| **Symbole** | **Abréviation de trois lettres** | **Nom de l’acide aminé** |
| A | Ala | Alanine |
| r | Arg | Arginine |
| n | Asn | Asparagine |
| d | Asp | Acide aspartique (aspartate) |
| c | Cys | Cystéine |
| E | Glu | Acide glutamique (glutamate) |
| Q | Gln | Glutamine |
| g | Gly | Glycine |
| h | His | Histidine |
| I | Ile | Isoleucine |
| L | Leu | Leucine |
| k | Lys | Lysine |
| m | Met | Méthionine |
| F | phe | Phénylalanine |
| P | Pro | Proline |
| O | Pyl | Pyrrolysine |
| s | Ser | Sérine |
| U | Sec | Sélénocystéine |
| t | Thr | Thréonine |
| w | Trp | Tryptophane |
| y | Tyr | Tyrosine |
| V | Val | Valine |
| b | Asx | Acide aspartique ou asparagine |
| Z | Glx | Glutamine ou acide glutamique |
| J | Xle | Leucine ou isoleucine |
| X | Xaa | A ou R ou N ou D ou C ou Q ou E ou G ou H ou I ou L ou K ou M ou F ou P ou O ou S ou U ou T ou W ou Y ou V; “unknown” ou “other” |

### Tableau B – Symboles, abréviations et noms conventionnels des acides aminés

## Index des exemples

**Page**

**Paragraphe 3.a) – Définition d’“acide aminé”**

**Exemple 3.a)-1 : Acides aminés D** [16](#page15)

***Exemples auxquels il est renvoyé***

Exemple 29-1 : Symbole ambigu le plus restrictif pour un acide aminé “other” [54](#page48)

Exemple 30-1 – Clé de caractérisation “CARBOHYD” [55](#page49)

**Paragraphe 3.c) – Définition de “énumeration de ses résidus”**

**Exemple 3.c)-1 : Énumération des acides aminés selon la structure chimique** [**17**](#example3c1)

**Exemple 3.c)-2 : Formule topologique pour une séquence d’acides aminés** [**18**](#example3c2)

***Exemples auxquels il est renvoyé***

Exemple 27-1 : Formule topologique pour une séquence de nucléotides [49](#page43)

Exemple 27-3 : Formule topologique – au moins quatre acides aminés définis de manière spécifique [51](#example273)

**Paragraphe 3.f) – Définition de “nucléotide modifié”**

***Exemples auxquels il est renvoyé***

Exemple 3.g)-4 : Analogues d’acides nucléiques [22](#example3g4)

**Paragraphe 3.g) – Définition de “nucléotide”**

**Exemple 3.g)-1 : Séquence de nucléotides interrompue par un espaceur C3** [**19**](#page18)

**Exemple 3.g)-2 : Séquence de nucléotides avec des résidus alternatifs, y compris un espaceur C3** [**20**](#example3g2)

**Exemple 3.g)-3 : Site abasique** [**21**](#example3g3)

**Exemple 3.g)-4 : Analogues d’acides nucléiques** [**22**](#example3g4)

Exemples auxquels il est renvoyé

Exemple 11.b)-1 : Séquence de nucléotides représentée par deux brins de codage – différentes longueurs [45](#page39)

Exemple 14-1 : Le symbole “t” représente l’uracile dans de l’ARN [47](#page41)

**Paragraphe 3.k) – Définition de “défini de manière spécifique”**

**Exemple 3.k)-1 : Symboles ambigus des nucléotides** [**23**](#page22)

**Exemple 3.k)-2 : Symbole ambigu “n” employé d’une manière à la fois conventionnelle et non conventionnelle** [**24**](#example3k2)

**Exemple 3.k)-3 : Symbole ambigu “n” employé d’une manière non conventionnelle** [**25**](#example3k3)

**Exemple 3.k)-4 : Les symboles ambigus autres que “n” sont “définis de manière spécifique”** [**26**](#example3k4)

**Exemple 3.k)-5 : Abréviation ambiguë “Xaa” employée d’une manière non conventionnelle** [**27**](#example3k5)

**Paragraphe 7 – Séquences devant être intégrées dans un listage**

***Exemples auxquels il est renvoyé***

Exemple 28-1: Codage de la séquence de nucléotides et séquence d’acides aminés codée [52](#page46)

Exemple 55-1 : Molécule combinée d’ADN et d’ARN [61](#Example551dnarna)

Exemple 87-2 : L’emplacement de la caractéristique s’étend au‑delà de la séquence divulguée [63](#Example872featurelocation)

Exemple 90-1 : Séquence d’acides aminés codée selon une séquence de codage avec introns [65](#page56)

**Paragraphe 7.a) – Séquences de nucléotides à intégrer dans un listage**

Exemple 7.a)-1 : Séquence de nucléotides ramifiée [**28**](#page27)

Exemple 7.a)-2 : Séquence de nucléotides linéaire comportant une structure secondaire [**30**](#example7a2linear)

Exemple 7.a)-3 : Symboles ambigus des nucléotides employés d’une manière non conventionnelle [**31**](#example7a3)

Exemple 7.a)-4 : Symboles ambigus des nucléotides employés d’une manière non conventionnelle [**32**](#example7a4)

Exemple 7.a)-5 : Symboles de nucléotides non conventionnels [**33**](#example7a5)

Exemple 7.a)-6 : Symboles de nucléotides non conventionnels [**34**](#example7a6)

Exemples auxquels il est renvoyé

Exemple 3.g)-1 : Séquence de nucléotides interrompue par un espaceur C3 [19](#page18)

Exemple 3.g)-2 : Séquence de nucléotides avec des résidus alternatifs, y compris un espaceur C3 [20](#example3g2)

Exemple 3.g)-3 : Site abasique [21](#example3g3)

Exemple 3.g)-4 : Analogues d’acides nucléiques [22](#example3g4)

Exemple 3.k)-1 : Symboles ambigus des nucléotides [23](#page22)

Exemple 3.k)-2 : Symbole ambigu “n” employé d’une manière à la fois conventionnelle et non conventionnelle [24](#example3k2)

Exemple 3.k)-3 : Symbole ambigu “n” employé d’une manière non conventionnelle [25](#example3k3)

Exemple 3.k)-4 : Les symboles ambigus autres que “n” sont “définis de manière spécifique” [26](#example3k4)

Exemple 11.a)-1 : Séquence de nucléotides représentée par deux brins de codage – mêmes longueurs [44](#page38)

Exemple 11.b)-1 : Séquence de nucléotides représentée par deux brins de codage – différentes longueurs [45](#page39)

Exemple 11.b)-2 : Séquence de nucléotides représentée par deux brins de codage – absence de segment d’appariement de bases [46](#example11b2)

Exemple 14-1 : Le symbole “t” représente l’uracile dans de l’ARN [47](#page41)

Exemple 87-1 : Codage de la séquence de nucléotides et séquence d’acides aminés codée [62](#page55)

Exemple 91-1 : Représentation des variantes énumérées [67](#page58)

Exemple 93.b)-1 : Représentation des séquences variantes individuelles comportant plusieurs variations interdépendantes [72](#page64)

**Paragraphe 7.b) – Séquences d’acides aminés à intégrer dans un listage**

**Exemple 7.b)-1 : Au moins quatre acides aminés définis de manière spécifique** [**35**](#page34)

**Exemple 7.b)-2 : Séquence d’acides aminés ramifiée** [**36**](#example7b2branched)

**Exemple 7.b)-3: Séquence d’acides aminés ramifiée** [**39**](#example7b3)

**Exemple 7.b)-4 : Peptide cyclique contenant une séquence d’acides aminés ramifiée** [**40**](#example7b4cyclicpeptide)

**Exemple 7.b)-5 : Peptide cyclique contenant une séquence d’acides aminés ramifiée ……………………………………**[**43**](#example7b5cyclicpeptide)

***Exemples auxquels il est renvoyé***

Exemple 3.a)-1 : Acides aminés D [16](#page15)

Exemple 3.c)-1 : Énumération des acides aminés selon la structure chimique [17](#example3c1)

Exemple 3.c)-2 : Formule topologique pour une séquence d’acides aminés [18](#example3c2)

Exemple 3.k)-5 : Abréviation ambiguë “Xaa” employée d’une manière non conventionnelle [27](#example3k5)

Exemple 27-1 : Formule topologique pour une séquence de nucléotides [49](#page43)

Exemple 27-3 : Formule topologique – au moins quatre acides aminés définis de manière spécifique [51](#example273)

Exemple 29-1 : Symbole ambigu le plus restrictif pour un acide aminé “other” [54](#page48)

Exemple 30-1 : Clé de caractérisation “CARBODHYD” [55](#page49)

Exemple 36-1 : Séquence dont une région contient un nombre connu de résidus “X” représentée comme une séquence unique [56](#example361)

Exemple 37-1 : Une séquence dont des régions contiennent un nombre inconnu de résidus “X” ne doit pas être représentée comme une séquence unique [59](#page53)

Exemple 37-2 : Une séquence dont des régions contiennent un nombre inconnu de résidus “X” ne doit pas être représentée comme une séquence unique [60](#example372)

Exemple 87-1 : Codage de la séquence de nucléotides et séquence d’acides aminés codée [62](#page55)

Exemple 91-2 : Représentation des variantes énumérées [68](#example912)

Exemple 91-3 : Représentation d’une séquence consensus [69](#example913)

Exemple 92-1 : Représentation d’une séquence unique avec des acides aminés alternatifs énumérés [70](#page62)

Exemple 93.a)-1 : Représentation d’une séquence variante par annotation de la séquence primaire [71](#page63)

**Paragraphe 8 – Seuil d’intégration des séquences**

***Exemples auxquels il est renvoyé***

Exemple 3.k)-1 : Symboles ambigus des nucléotides [23](#page22)

Exemple 3.k)-2 : Symbole ambigu “n” employé d’une manière à la fois conventionnelle et non conventionnelle [24](#example3k2)

Exemple 7.a)-1 : Séquence de nucléotides ramifiée [28](#page27)

Exemple 7.a)-6 : Symboles de nucléotides non conventionnels [34](#example7a6)

Exemple 7.b)-1: Au moins quatre acides aminés définis de manière spécifique [35](#page34)

Exemple 7.b)-2 : Séquence d’acides aminés ramifiée [36](#example7b2branched)

Exemple 7.b)-4 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [40](#example7b4cyclicpeptide)

Exemple 14-1 : Le symbole “t” représente l’uracile dans de l’ARN [47](#page41)

Exemple 37-1 : Une séquence dont des régions contiennent un nombre inconnu de résidus “X” ne doit pas être représentée comme une séquence unique [59](#page53)

Exemple 37-2 : Une séquence dont des régions contiennent un nombre inconnu de résidus “X” ne doit pas être représentée comme une séquence unique [60](#example372)

Exemple 70-1: Représentation d’une séquence unique avec des acides aminés alternatifs énumérés [70](#page62)

**Paragraphe 11 – Représentation d’une séquence de nucléotides**

**Exemples auxquels il est renvoyé**

Exemple 3.g)-4 : Analogues d’acides nucléiques [22](#example3g4)

Exemple 7.a)-1 : Séquence de nucléotides ramifiée [28](#page27)

**Paragraphe 11.a) – Séquence de nucléotides représentée par deux brins de codage – entièrement complémentaires**

**Exemple 11.a)-1 : Séquence de nucléotides représentée par deux brins de codage – mêmes longueurs** [**44**](#page38)

**Paragraphe 11.b) – Séquence de nucléotides représentée par deux brins de codage – non entièrement complémentaires**

**Exemple 11.b)-1 : Séquence de nucléotides représentée par deux brins de codage – différentes longueurs** [**45**](#page39)

**Exemple 11.b)-2 : Séquence de nucléotides représentée par deux brins de codage – absence de segment d’appariement de** **bases** [**46**](#example11b2)

**Paragraphe 13 – Représentation des nucléotides**

***Exemples auxquels il est renvoyé***

Exemple 3.k)-2 : Symbole ambigu “n” employé d’une manière à la fois conventionnelle et non conventionnelle [24](#example3k2)

Exemple 7.a)-1 : Séquence de nucléotides ramifiée [28](#page27)

Exemple 14-1 : Le symbole “t” représente l’uracile dans de l’ARN [47](#page41)

Exemple 91-1 : Représentation des variantes énumérées [67](#page58)

**Paragraphe 14 – Le symbole “t”désigne l’uracile dans de l’ARN**

**Exemple 14-1: Le symbole “t” représente l’uracile dans de l’ARN** [**47**](#page41)

***Exemples auxquels il est renvoyé***

Exemple 55-1 : Molécule combinée d’ADN et d’ARN [61](#Example551dnarna)

**Paragraphe 15 – Il faut choisir le symbole ambigu des nucléotides le plus restrictif**

***Exemples auxquels il est renvoyé***

Exemple 3.g)-1 : Séquence de nucléotides interrompue par un espaceur C3 [19](#page18)

Exemple 3.g)-2 : Séquence de nucléotides avec des résidus alternatifs, y compris un espaceur C3 [20](#example3g2)

Exemple 3.k)-4 : Les symboles ambigus autres que “n” sont “définis de manière spécifique” [26](#example3k4)

Exemple 93.b)-1 : Représentation des séquences variantes individuelles comportant plusieurs variations interdépendantes [72](#page64)

**Paragraphe 16 – Représentation d’un acide aminé modifié**

***Exemples auxquels il est renvoyé***

Exemple 3.g)-1 : Séquence de nucléotides interrompue par un espaceur C3 [19](#page18)

Exemple 3.g)-4 : Analogues d’acides nucléiques [22](#example3g4)

**Paragraphe 17 – Annotation d’un acide aminé modifié**

***Exemples auxquels il est renvoyé***

Exemple 3.g)-1 : Séquence de nucléotides interrompue par un espaceur C3 [19](#page18)

Exemple 3.g)-3: Site abasique [21](#example3g3)

Exemple 7.a)-1: Séquence de nucléotides ramifiée [28](#page27)

Exemple 7.a)-2 : Séquence de nucléotides linéaire comportant une structure secondaire [30](#example7a2linear)

Exemple 7.a)-6: Symboles de nucléotides non conventionnels [34](#example7a6)

**Paragraphe 18 – Annotation des régions de nucléotides modifiés consécutifs**

***Exemples auxquels il est renvoyé***

Exemple 3.g)-4 : Analogues d’acides nucléiques [22](#example3g4)

Exemple 11.b)-1 : Séquence de nucléotides représentée par deux brins de codage – différentes longueurs [45](#page39)

**Paragraphe 19 – Annotation de l’uracile dans de l’ADN ou de la thymine dans de l’ARN**

**Exemples auxquels il est renvoyé**

Exemple 14-1: Le symbole “t” représente l’uracile dans de l’ARN [47](#page41)

**Paragraphe 25 – Position de résidu numéro 1 de la séquence d’acides aminés**

***Exemples auxquels il est renvoyé***

Exemple 3.a)-1: Acides aminés D [16](#page15)

Exemple 7.b)-4 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [40](#example7b4cyclicpeptide)

Exemple 7.b)-5 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [43](#example7b5cyclicpeptide)

Exemple 29-1 : Symbole ambigu le plus restrictif pour un acide aminé “other” [54](#page48)

**Paragraphe 26 – Représentation des acides aminés**

***Exemples auxquels il est renvoyé***

Exemple 7.b)-2 : Séquence d’acides aminés ramifiée [36](#example7b2branched)

Exemple 7.b)-4 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [40](#example7b4cyclicpeptide)

Exemple 7.b)-5 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [43](#example7b5cyclicpeptide)

Exemple 36-1 : Séquence dont une région contient un nombre connu de résidus “X” représentée comme une séquence unique [56](#example361)

Exemple 87-1 : Codage de la séquence de nucléotides et séquence d’acides aminés codée [62](#page55)

Exemple 90-1 : Séquence d’acides aminés codée selon une séquence de codage avec introns [65](#page56)

Exemple 91-2 : Représentation des variantes énumérées [68](#example912)

Exemple 91-3 : Représentation d’une séquence consensus [69](#example913)

**Paragraphe 27 – Il faut choisir le symbole ambigu des acides aminés le plus restrictif**

**Exemple 27-1 : Formule topologique pour une séquence de nucléotides** [**49**](#page43)

**Exemple 27-2 : Formule topologique – moins de quatre acides aminés définis de manière spécifique** [**50**](#example272)

**Exemple 27-3 : Formule topologique – au moins quatre acides aminés définis de manière spécifique** [**51**](#example273)

***Exemples auxquels il est renvoyé***

Exemple 3.c)-2 : Formule topologique pour une séquence d’acides aminés [18](#example3c2)

Exemple 7.b)-1 : Au moins quatre acides aminés définis de manière spécifique [35](#page34)

Exemple 29-1 : Symbole ambigu le plus restrictif pour un acide aminé “other” [54](#page48)

Exemple 36-1 : Séquence dont une région contient un nombre connu de résidus “X” représentée comme une séquence unique [56](#example361)

Exemple 36-2 : Séquence dont plusieurs régions contiennent un nombre ou une série connu de résidus “X” représentée comme une séquence unique [57](#example362)

Exemple 36-3 : Séquence dont plusieurs régions contiennent un nombre ou une série connu de résidus “X” représentée comme une séquence unique [58](#example363)

Exemple 37-2 : Une séquence dont des régions contiennent un nombre inconnu de résidus “X” ne doit pas être représentée comme une séquence unique [60](#example372)

Exemple 91-3 : Représentation d’une séquence consensus [69](#example913)

Exemple 92-1 : Représentation d’une séquence unique avec des acides aminés alternatifs énumérés [70](#page62)

Exemple 93.a)-1 : Représentation d’une séquence variante par annotation de la séquence primaire [71](#page63)

**Paragraphe 28 – Séquences d’acides aminés séparées par des symboles internes de fin**

**Exemple 28-1 : Codage de la séquence de nucléotides et séquence d’acides aminés codée** [**52**](#page46)

***Exemples auxquels il est renvoyé***

Exemple 87-1 : Codage de la séquence de nucléotides et séquence d’acides aminés codée [62](#page55)

Exemple 90-1 : Séquence d’acides aminés codée selon une séquence de codage avec introns [65](#page56)

**Paragraphe 29 – Représentation d’un acide aminé modifié “other”**

**Exemple 29-1 : Symbole ambigu le plus restrictif pour un acide aminé “other”** [**54**](#page48)

***Exemples auxquels il est renvoyé***

Exemple 3.a)-1: Acides aminés D [16](#page15)

Exemple 7.b)-2: Séquence d’acides aminés ramifiée [36](#example7b2branched)

Exemple 7.b)-3 : Séquence d’acides aminés ramifiée [39](#example7b3)

Exemple 7.b)-4 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [40](#example7b4cyclicpeptide)

Exemple 30-1 : Clé de caractérisation “CARBODHYD” [55](#page49)

**Paragraphe 30 – Annotation d’un acide aminé modifié**

**Exemple 30-1 : Clé de caractérisation “CARBODHYD”** [**55**](#page49)

***Exemples auxquels il est renvoyé***

Exemple 3.a)-1 : Acides aminés D [16](#page15)

Exemple 3.c)-1 : Énumération des acides aminés selon la structure chimique [17](#example3c1)

Exemple 7.b)-2 : Séquence d’acides aminés ramifiée [36](#example7b2branched)

Exemple 7.b)-3: Séquence d’acides aminés ramifiée [39](#example7b3)

Exemple 7.b)-4 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [40](#example7b4cyclicpeptide)

Exemple 7.b)-5 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [43](#example7b5cyclicpeptide)

Exemple 29-1 : Symbole ambigu le plus restrictif pour un acide aminé “other” [54](#example291)

**Paragraphe 31 – Représentation d’un acide aminé D**

***Exemples auxquels il est renvoyé***

Exemple 3.a)-1 : Acides aminés D [16](#page15)

Exemple 3.c)-1 : Énumération des acides aminés selon la structure chimique [17](#example3c1)

Exemple 7.b)-2 : Séquence d’acides aminés ramifiée [36](#example7b2branched)

Exemple 7.b)-3: Séquence d’acides aminés ramifiée [39](#example7b3)

Exemple 7.b)-4 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [40](#example7b4cyclicpeptide)

Exemple 7.b)-5 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [43](#example7b5cyclicpeptide)

**Paragraphe 32 – Annotation d’un acide aminé “unknown”**

***Exemples auxquels il est renvoyé***

Exemple 3.c)-1 : Énumération des acides aminés selon la structure chimique [17](#example3c1)

**Paragraphe 34 – Annotation d’une région contiguë de résidus “X”**

***Exemples auxquels il est renvoyé***

Exemple 29-1 : Symbole ambigu le plus restrictif pour un acide aminé “other” [54](#page48)

**Paragraphe 36 – Séquences contenant des régions comportant un nombre exact de résidus contigus “n” ou “X”**

**Exemple 36-1 : Séquence dont une région contient un nombre connu de résidus “X” représentée comme une séquence unique** [**56**](#example361)

**Exemple 36-2 : Séquence dont plusieurs régions contiennent un nombre ou une série connu de résidus “X” représentée comme une séquence unique** [**57**](#example362)

**Exemple 36-3 : Séquence dont plusieurs régions contiennent un nombre ou une série connu de résidus “X” représentée comme une séquence unique** [**58**](#example363)

**Paragraphe 37 – Séquences dont des régions contiennent un nombre inconnu de résidus contigus “n” ou “X”**

**Exemple 37-1 : Une séquence dont des régions contiennent un nombre inconnu de résidus “X” ne doit pas être représentée comme une séquence unique** [**59**](#page53)

**Exemple 37-2 : Une séquence dont des régions contiennent un nombre inconnu de résidus “X” ne doit pas être représentée comme une séquence unique** [**60**](#example372)

**Paragraphe 41 – Caractères réservés**

***Exemples auxquels il est renvoyé***

Exemple 87-2 : L’emplacement de la caractéristique s’étend au‑delà de la séquence divulguée [63](#Example872featurelocation)

**Paragraphe 54 – L’élément INSDSeq\_moltype**

***Exemples auxquels il est renvoyé***

Exemple 14-1: Le symbole “t” représente l’uracile dans de l’ARN [47](#page41)

**Paragraphe 55 – Séquence de nucléotides contenant des segments d’ADN et d’ARN**

**Exemple 55-1: Molécule combinée d’ADN et d’ARN** [61](#Example551dnarna)

**Paragraphe 56 – Exemple illustrant une séquence nucléotidique contenant des segments d’ADN et d’ARN**

***Exemples auxquels il est renvoyé***

Exemple 55-1 : Molécule combinée d’ADN et d’ARN[61](#Example551dnarna)

**Paragraphe 57 – L’élément INSDSeq\_sequence**

***Exemples auxquels il est renvoyé***

Exemple 28-1 : Codage de la séquence de nucléotides et séquence d’acides aminés codée [52](#page46)

Exemple 90-1 : Séquence d’acides aminés codée selon une séquence de codage avec introns [65](#page56)

**Paragraphe 65 – Descripteur d’emplacement**

***Exemples auxquels il est renvoyé***

Exemple 3.g)-4: Analogues d’acides nucléiques [22](#example3g4)

Exemple 87-2 : L’emplacement de la caractéristique s’étend au‑delà de la séquence divulguée [63](#Example872featurelocation)

***Paragraphe 66 – Syntaxe des descripteurs d’emplacement***

***Exemples auxquels il est renvoyé***

Exemple 3.g)-4 : Analogues d’acides nucléiques [22](#example3g4)

Exemple 29-1: Symbole ambigu le plus restrictif pour un acide aminé “other” [54](#page48)

Exemple 87-2: L’emplacement de la caractéristique s’étend au‑delà de la séquence divulguée [63](#Example872featurelocation)

**Paragraphe 67 – Opérateur d’emplacement**

***Exemples auxquels il est renvoyé***

Exemple 7.b)-4 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [40](#example7b4cyclicpeptide)

**Paragraphe 68 – Opérateurs d’emplacement “join” et “order”**

***Exemples auxquels il est renvoyé***

Exemple 7.b)-4 : Peptide cyclique contenant une séquence d’acides aminés ramifiée [40](#example7b4cyclicpeptide)

**Paragraphe 70 – Emplacements de caractéristiques**

**Exemples auxquels il est renvoyé**

Exemple 29-1: Symbole ambigu le plus restrictif pour un acide aminé “other” [54](#page48)

Exemple 87-2 : L’emplacement de la caractéristique s’étend au‑delà de la séquence divulguée [63](#Example872featurelocation)

**Paragraphe 71 – Représentation des caractères “<” et “>” dans un descripteur d’emplacement**

**Exemples auxquels il est renvoyé**

Exemple 29-1: Symbole ambigu le plus restrictif pour un acide aminé “other” [54](#page48)

Exemple 87-2 : L’emplacement de la caractéristique s’étend au‑delà de la séquence divulguée [63](#Example872featurelocation)

**Paragraphe 83 – Exemple illustrant une séquence de nucléotides n’existant pas à l’état naturel**

***Exemples auxquels il est renvoyé***

Exemple 55-1 : Molécule combinée d’ADN et d’ARN [61](#Example551dnarna)

**Paragraphe 87 – Clé de caractérisation “CDS”**

Exemple 87-1: Codage de la séquence de nucléotides et séquence d’acides aminés codée [**62**](#page55)

**Exemple 87-2 : L’emplacement de la caractéristique s’étend au‑delà de la séquence divulguée** [**63**](#Example872featurelocation)

***Exemples auxquels il est renvoyé***

Exemple 90-1 : Séquence d’acides aminés codée selon une séquence de codage avec introns [65](#page56)

**Paragraphe 88 – Les qualificateurs “transl\_table” et “translation”**

***Exemples auxquels il est renvoyé***

Exemple 28-1 : Codage de la séquence de nucléotides et séquence d’acides aminés codée [52](#page46)

Exemple 87-1 : Codage de la séquence de nucléotides et séquence d’acides aminés codée [62](#page55)

Exemple 90-1 : Séquence d’acides aminés codée selon une séquence de codage avec introns [65](#page56)

**Paragraphe 90 – Intégration d’une séquence d’acides aminés codée dans un listage des séquences**

**Exemple 90-1 : Séquence d’acides aminés codée selon une séquence de codage avec introns** [**65**](#page56)

***Exemples auxquels il est renvoyé***

Exemple 28-1 : Codage de la séquence de nucléotides et séquence d’acides aminés codée [52](#page46)

Exemple 87-1: Codage de la séquence de nucléotides et séquence d’acides aminés codée [62](#page55)

Exemple 87-2 : L’emplacement de la caractéristique s’étend au‑delà de la séquence divulguée[63](#Example872featurelocation)

**Paragraphe 91 – Séquence primaire et une variante, chacune énumérée par son résidu**

**Exemple 91-1 : Représentation des variantes énumérées** [**67**](#page58)

**Exemple 91-2 : Représentation des variantes énumérées** [**68**](#example912)

Exemple 91-3 : Représentation d’une séquence consensus [**69**](#example913)

**Paragraphe 92 – Séquence variante divulguée comme une séquence unique avec des résidus alternatifs énumérés**

**Exemple 92-1 : Représentation d’une séquence unique avec des acides aminés alternatifs énumérés** [**70**](#page62)

**Paragraphe 93.a) – Toute séquence variante divulguée uniquement par référence à une séquence primaire comportant plusieurs variations indépendantes**

**Exemple 93.a)-1 : Représentation d’une séquence variante par annotation de la séquence primaire** [**71**](#page63)

**Paragraphe 93.b) – Toute séquence variante divulguée uniquement par référence à une séquence primaire comportant plusieurs variations interdépendantes**

**Exemple 93.b)-1 : Représentation des séquences variantes individuelles comportant plusieurs variations interdépendantes** [**72**](#page64)

**Paragraphe 94 – Clés de caractérisation et qualificateurs pour une séquence variante**

***Exemples auxquels il est renvoyé***

Exemple 29-1 : Symbole ambigu le plus restrictif pour un acide aminé “other” [54](#page48)

**Paragraphe 95– Annotation d’une séquence variante**

***Exemples auxquels il est renvoyé***

Exemple 29-1 : Symbole ambigu le plus restrictif pour un acide aminé “other” [54](#page48)

Exemple 91-3 : Représentation d’une séquence consensus [69](#example913)

Exemple 92-1 : Représentation d’une séquence unique avec des acides aminés alternatifs énumérés [70](#page62)

## Exemples

### Paragraphe 3.a) Définition d’“acide aminé”

**Exemple 3.a)-1 : acides aminés D**

Une demande de brevet décrit la séquence ci‑après :

Cyclo (D-Ala-D-Glu-Lys-Nle-Gly-D-Met-D-Nle)

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Le paragraphe 3.a) de la norme définit “acide aminé” comme comprenant les “acides aminés D” et les acides aminés contenant des chaînes latérales modifiées ou synthétiques. Conformément à cette définition, le peptide énuméré contient cinq acides aminés qui sont définis de manière spécifique (D-Ala, D-Glu, Lys, Gly et D-Met). La séquence doit donc être intégrée dans un listage des séquences comme l’exige le paragraphe 7.b) de la norme ST.26.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Le paragraphe 29 prescrit que les acides aminés modifiés devraient être représentés dans la séquence comme l’acide aminé L non modifié correspondant. Par ailleurs, tout acide aminé modifié ne pouvant être représenté à l’aide d’un autre symbole indiqué dans l’annexe I, section 3, tableau 3 doit être représenté par le symbole “X”.

Dans cet exemple, la séquence contient trois acides aminés D qui peuvent être représentés par un acide aminé L non modifié indiqué dans l’annexe I, section 3, tableau 3, un acide aminé L (Nle) et un acide aminé D (D-Nle) qui doit être représenté par le symbole “X”.

Aux termes du paragraphe 25, lorsque les séquences d’acides aminés présentent une configuration circulaire et que l’anneau se compose uniquement de résidus d’acides aminés ~~dans des liaisons~~liés par des peptides, le déposant doit choisir l’acide aminé à la position de résidu numéro 1. La séquence peut donc être représentée comme suit :

AEKXGMX (SEQ ID NO : 1)

ou autrement, avec un autre acide aminé de la séquence en position de résidu numéro 1. Une clé de caractérisation “SITE” et un qualificateur “NOTE” doivent être indiqués pour chaque acide aminé D, le nom complet non abrégé de l’acide aminé D étant indiqué dans la valeur du qualificateur, par exemple l’alanine ~~A~~D et la norleucine ~~N~~D. Par ailleurs, une clé de caractérisation “SITE” et un qualificateur “NOTE” doivent être indiqués, le qualificateur prenant comme valeur l’abréviation de la norleucine L, c’est‑à‑dire “Nle”, conformément à ce qui est indiqué dans l’annexe I, section 4, tableau 4. Enfin, une clé de caractérisation “REGION” et un qualificateur “NOTE” devraient être prévus pour montrer que le peptide est circulaire.

**Paragraphes pertinents de la norme ST.26 :** ~~paragraphes~~**3.a)**, 7.b), 25, 26, 29, 30 et 31

### Paragraphe 3.c) – Définition de “énumeration de ses résidus”

**Exemple 3.c)-1 : Énumération des acides aminés selon la structure chimique**

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Le peptide énuméré, illustré sous forme d’une structure, contient au moins quatre acides aminés définis de manière spécifique. La séquence doit donc être intégrée dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence peut être représentée sous la forme suivante :

VAFXGK (SEQ ID NO : 2)

où “X” représente un acide aminé modifié “other”, qui nécessité une clé de caractérisation “SITE” ainsi que le qualificateur “NOTE”. Le qualificateur “NOTE” indique le nom complet non abrégé du tryptophane modifié à la position 4 du peptide énuméré, par exemple l’“acide 6-amino-7-(1H-indol-3-yl)-5-oxoheptanoïque”. D’autre part, des clés de caractérisation “SITE” et un qualificateur “NOTE” supplémentaires sont requis pour indiquer l’acétylation de l’extrémité N-terminale et la méthylation de l’extrémité C-terminale.

La séquence pourrait également être représentée sous la forme suivante :

VAFW (SEQ ID NO : 3)

Une clé de caractérisation “SITE” et un qualificateur “NOTE” sont requis pour indiquer une modification du tryptophane à la position 4 du peptide énuméré avec la valeur “extrémité C-terminale liée au dipeptide GK par l’intermédiaire d’un pont de glutaraldéhyde”. D’autre part, une clé de caractérisation “SITE” à l’emplacement 1 et un qualificateur “NOTE” supplémentaires sont requis pour indiquer l’acétylation de l’extrémité N-terminale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**3.c)**, 7.b), 26, 29, 30 et 31

**Exemple 3.c)-2 : Formule topologique pour une séquence d’acides aminés**

(G4z)n

où G= Glycine, z = tout acide aminé et la variable n peut être tout nombre entier.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La divulgation indique que “n” peut être “tout nombre entier”; le mode de réalisation le plus englobant de “n” est donc indéterminé. Étant donné que “n” est indéterminé, le peptide de la formule ne peut pas être amplifié jusqu’à une longueur déterminée et, de ce fait, la formule non amplifiée doit être considérée.

Le peptide énuméré de la formule non amplifiée (“n” = 1) contient quatre acides aminés définis de manière spécifique, dont chacun est Gly, et le symbole “z”. Par convention, “Z” est le symbole de la “glutamine ou (de l’) acide glutamique”; toutefois, l’exemple définit “z” comme “tout acide aminé”. Conformément à la norme ST.26, un acide aminé qui n’est pas défini de manière spécifique est représenté par “X”. Based on this analysis, the enumerated peptide, i.e., GGGGX, contains four glycine residues that are enumerated and specifically defined. De ce fait, le paragraphe 7.b) de la norme ST.26 prescrit l’intégration de la séquence dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence emploie un symbole non conventionnel “z”, dont la définition doit être déterminée à partir de la divulgation (voir l’Introduction au présent document). Étant donné que “z” est défini comme tout acide aminé, le symbole conventionnel servant à représenter cet acide aminé est “X.” De ce fait, la séquence doit être représentée comme une séquence unique :

GGGGX (SEQ ID NO : 4)

de préférence annotée à l’aide de la clé de caractérisation REGION, de l’emplacement de caractéristique “&gt;5” (correspond à >5) et d’un qualificateur NOTE qui prend comme valeur “L’ensemble de la séquence d’acides aminés 1-5 peut être répété une ou plusieurs fois”.

Conformément au paragraphe 27, “X” sera considéré comme l’équivalent de l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire dans le tableau de caractéristiques. Dès lors, si “X” est destiné à représenter “tout acide aminé”, il doit être annoté au moyen de la clé de caractérisation “VARIANT” et d’un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

Chaque fois que possible, chaque “X” devrait être annoté individuellement. Cependant, une région de résidus “X” contigus ou un grand nombre de résidus “X” dispersés dans l’ensemble de la séquence peuvent être décrits conjointement par la clé de caractérisation “VARIANT” en employant la syntaxe “x.y” pour désigner le descripteur d’emplacement, où x et y sont les positions du premier et du dernier résidu “X”, et par un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

**N.B. :** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**3.c)** ~~et~~, 7.b) et 27

### Paragraphe 3.g) Définition de “nucléotide”

**Exemple 3.g)-1 : Séquence de nucléotides interrompue par un espaceur C3**

Une demande de brevet décrit la séquence ci‑après :

atgcatgcatgcncggcatgcatgc

où n = un espaceur C3 dont la structure est la suivante :



**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La séquence énumérée contient deux segments de nucléotides définis de manière spécifique et séparés par un espaceur C3.

L’espaceur C3 n’est pas un nucléotide au sens du paragraphe 3.g); le symbole conventionnel “n” est employé d’une manière non conventionnelle (voir l’Introduction au présent document). En conséquence, chaque segment est une séquence de nucléotides distincte. Étant donné que chaque segment contient plus de 10 nucléotides définis de manière spécifique, les deux segments doivent être intégrés dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Chaque segment doit être intégré dans un listage des séquences en tant que séquence distincte, chaque séquence devant disposer de son propre numéro d’identification de séquence :

atgcatgcatgc (SEQ ID NO : 5)

cggcatgcatgc (SEQ ID NO : 6)

La cytosine contenue dans chaque segment attaché à l’espaceur C3 devrait être accompagnée d’une description supplémentaire dans un tableau de caractéristiques en employant la clé de caractérisation “misc\_feature” et le qualificateur “note”. La valeur du qualificateur “note”, qui est “free text”, devrait indiquer la présence de l’espaceur, qui est lié à un autre acide nucléique, et devrait identifier l’espaceur soit par son nom chimique complet non abrégé, soit par son nom commun, par exemple espaceur C3.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**3.g)**, 7.a) et 15

**Exemple 3.g)-2 : Séquence de nucléotides avec des résidus alternatifs, y compris un espaceur C3**

Une demande de brevet décrit la séquence ci‑après :

atgcatgcatgcncggcatgcatgc

où n = c, a, g ou un espaceur C3 dont la structure est la suivante :



**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Il y a 24 résidus définis de manière spécifique dans la séquence énumérée interrompue par la variable “n.” L’explication de la séquence figurant dans la divulgation doit être consultée pour déterminer si le “n” est employé d’une manière conventionnelle ou non conventionnelle (voir l’Introduction au présent document).

La divulgation indique n = c, a, g ou un espaceur C3. Le “n” est un symbole conventionnel employé d’une manière non conventionnelle, puisqu’il est décrit comme englobant un espaceur C3, qui ne répond pas à la définition d’un nucléotide. Le symbole “n” est également décrit comme englobant “c”, “a” ou “g”; il s’ensuit que la norme ST.26 prescrit l’intégration de la séquence des 25 nucléotides dans un listage des séquences. Étant donné que deux segments séparés par l’espaceur C3 sont des séquences distinctes de la séquence des 25 nucléotides, les deux séquences de 12 nucléotides peuvent également être intégrées.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

L’exemple indique que “n = c, a, g ou un espaceur C3”. Comme on l’a vu plus haut, un espaceur C3 n’est pas un nucléotide. Conformément au paragraphe 15, le symbole “n” ne doit être employé que pour représenter un nucléotide; il s’ensuit que le symbole “n” ne peut pas représenter un espaceur C3 dans un listage des séquences.

Le paragraphe 15 dit également que, lorsqu’il convient d’employer un symbole ambigu, il faudrait choisir le symbole le plus restrictif. Conformément à l’annexe I, section 1, tableau 1, le symbole “v” représente “a or c ou g”, qui est plus restrictif que “n”.

Lorsque, dans l’exemple, la variable “n” est c, a ou g, la séquence unique énumérée par ses résidus qui englobe les modes de réalisation les plus divulgués et est, de ce fait, la séquence la plus englobante (voir l’Introduction au présent document) à intégrer dans un listage des séquences est la suivante :

atgcatgcatgcvcggcatgcatgc (SEQ ID NO : 7)

Comme indiqué dans l’introduction au présent document, il est fortement conseillé d’intégrer toutes séquences supplémentaires de première importance pour la divulgation ou les revendications de l’invention.

Lorsque, dans l’exemple, la variable “n” est un espaceur C3, la séquence peut être considérée comme deux segments distincts de nucléotides définis de manière spécifique de chaque côté de la variable “n”, c’est‑à‑dire atgcatgcatgc (SEQ ID NO : 8) et cggcatgcatgc (SEQ ID NO : 9). Si elles sont de première importance pour la divulgation ou les revendications, ces deux séquences devraient également être intégrées dans le listage des séquences, chacune disposant de son propre numéro d’identification de séquence.

La cytosine contenue dans chaque segment attaché à l’espaceur C3 devrait être accompagnée d’une description supplémentaire dans un tableau de caractéristiques en employant la clé de caractérisation “misc\_feature” et le qualificateur “note”. The “note” qualifier value, which is “free text”, should indicate the presence of the spacer, which is joined to another nucleic acid and identify the spacer by either its complete unabbreviated chemical name, or by its common name, e.g., C3 spacer.

**N.B. :** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

Paragraphe(s) pertinent(s) de la norme ST.26 : ~~paragraphes~~**3.g)**, 7.a) et 15

**Exemple 3.g)-3 : Site abasique**

Une demande de brevet décrit la séquence ci‑après :

gagcattgac-AP-taaggct

où AP est un site abasique

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Les résidus définis d’une manière spécifique de la séquence énumérée sont interrompus par un site abasique. Le côté 5’ du site abasique contient 10 nucléotides et son côté 3’ en contient sept. Le paragraphe 3.g)ii)2) définit un site abasique comme un “nucléotide” lorsqu’il fait partie d’une séquence nucléotidique. Il s’ensuit que, dans cet exemple, le site abasique est considéré comme un “nucléotide” pour ce qui est de déterminer si et comment la séquence doit être intégrée dans un listage des séquences. En conséquence, les résidus se trouvant de chaque côté du site abasique font partie d’une séquence énumérée unique contenant au total 18 nucléotides, dont 17 sont définis de manière spécifique. La séquence doit donc être intégrée en tant que séquence unique comme le prescrit le paragraphe 7.~~b~~a) de la norme ST.26.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence doit être intégrée dans un listage des séquences sous la forme suivante :

gagcattgacntaaggct (SEQ ID NO : 10)

Le site abasique doit être représenté par un “n” et doit être accompagné d’une description supplémentaire dans un tableau de caractéristiques. Le moyen d’annotation à préférer est la clé de caractérisation “modified\_base” et le qualificateur obligatoire est le “mod\_base” qui prend la valeur “OTHER”. Il faut intégrer un qualificateur “note”, qui décrive la base modifiée comme étant un site abasique.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**3.g)**, 7.a) et 17

**Exemple 3.g)-4 : Analogues d’acides nucléiques**

Une demande de brevet divulgue la séquence d’un acide nucléique à glycol (ANG) suivante :

PO4-tagttcattgactaaggctccccattgact-OH

où l’extrémité gauche de la séquence reproduit le côté 5’ d’une séquence d’ADN.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

OUI – Les résidus individuels qui constituent une séquence d’ANG sont considérés comme des nucléotides en vertu du paragraphe 3.g)i)2) de la norme ST.26. Il s’ensuit que la séquence compte plus de 10 nucléotides énumérés et “définis de manière spécifique” et doit être intégrée dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Les séquences d’ANG n’ont pas d’extrémité 5’ ni d’extrémité 3’, mais une extrémité 3’ et une extrémité 2’. L’extrémité 3’, qui est normalement décrite comme ayant un groupe phosphate terminal, correspond à l’extrémité 5’ de l’ADN ou de l’ARN. (A noter que d’autres analogues d’acides nucléiques peuvent correspondre de façon différente à l’extrémité 5’ et à l’extrémité 3’ de l’ADN et de l’ARN.) En vertu du paragraphe ~~10~~11, la séquence doit être intégrée dans un listage des séquences “de gauche à droite de manière à reproduire le sens 5’-3’”. Elle doit donc être intégrée dans un listage des séquences sous la forme suivante :

tagttcattgactaaggctccccattgact (SEQ ID NO : 11)

La séquence doit être décrite dans le tableau de caractéristiques à l’aide de la clé de caractérisation “modified\_base” et du qualificateur obligatoire “mod\_base” avec l’abréviation “OTHER”. Un qualificateur de type “note” doit être intégré avec le nom complet non abrégé des nucléotides modifiés, par exemple les “acides nucléiques à glycol ou les “2,3-dihydroxypropyl nuclosides”. Un élément INSDFeature unique peut être employé pour décrire l’ensemble de la séquence comme un ANG où l’élément INSDFeature\_location présente la série “1.30”.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~3.d), **3.g)**, 7.a), 11, 16, 18, 65 et 66

### Paragraphe 3.k) Définition de “défini de manière spécifique”

**Exemple 3.k)-1 : Symboles ambigus des nucléotides**

5’ NNG KNG KNG K 3’

**N et K sont des codes ambigus IUPAC-IUB**

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**NON**

Les codes ambigus IUPAC-IUB correspondent à la liste des symboles des nucléotides définis à l’annexe I, section 1, tableau 1. Selon le paragraphe 3.k), un nucléotide défini de manière spécifique est un nucléotide différent de ceux qui sont représentés par le symbole “n” et énumérés à l’annexe I. En conséquence, “K” et “G” sont des nucléotides définis de manière spécifique tandis que “N” n’en est pas un.

La séquence énumérée n’ayant pas au moins 10 nucléotides définis de manière spécifique, le paragraphe 7.a) de la norme ST.26 ne prescrit pas son intégration dans un listage des séquences.

**Question 2 : La norme ST.26 autorise‑t‑elle l’intégration de la ou des séquences?**

**NON**

En vertu du paragraphe 8, “un listage des séquences ne doit contenir aucune séquence comportant moins de 10 nucléotides définis de manière spécifique…”. N’ayant pas au moins 10 nucléotides définis de manière spécifique, la séquence énumérée ne doit pas être intégrée dans un listage des séquences.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**3.k)**, 7.a), 8 et 13

**Exemple 3.k)-2 : Symbole ambigu “n” employé d’une manière à la fois conventionnelle et non conventionnelle**

Une demande divulgue la séquence artificielle suivante : 5’-AATGCCGGAN-3’. La divulgation indique également ce qui suit :

i) dans un mode de réalisation, N représente tout nucléotide;

ii) dans un mode de réalisation, N est facultatif mais représente de préférence G;

iii) dans un mode de réalisation, N représente K;

iv) dans un mode de réalisation, N représente C.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**NON**

La séquence énumérée contient neuf nucléotides définis de manière spécifique et un “N.” Il faut consulter l’explication de la séquence donnée dans la divulgation afin de déterminer si le symbole “N” est utilisé d’une manière conventionnelle (voir l’Introduction au présent document).

L’examen des modes de réalisation divulgués i) à iv) de la séquence énumérée montre que le mode de réalisation de “N” le plus englobant est “tout nucléotide”. Dans le mode de réalisation le plus englobant, “N” est utilisé d’une manière conventionnelle dans la séquence énumérée.

Pour certains modes de réalisation, “N” est décrit comme représentant des résidus définis de manière spécifique (c’est‑à‑dire “N représente C” dans la partie iv)). Toutefois, seul le mode de réalisation le plus englobant (c’est‑à‑dire “N représente tout nucléotide”) est pris en considération lorsqu’il s’agit de déterminer si une séquence doit être intégrée dans un listage des séquences. Il s’ensuit que la séquence énumérée qui doit être évaluée est 5’-AATGCCGGAN-3’.

Sur la base de cette analyse, la séquence énumérée, à savoir AATGCCGGAN, ne contient pas 10 nucléotides définis de manière spécifique. De ce fait, le paragraphe 7.a) de la norme ST.26 ne prescrit pas l’intégration de la séquence dans un listage des séquences, en dépit du fait que “n” est également défini comme des nucléotides particuliers dans certains modes de réalisation.

**Question 2 : La norme ST.26 autorise‑t‑elle l’intégration de la ou des séquences?**

**NON**

La séquence “AATGCCGGAN” ne doit pas être intégrée dans un listage des séquences.

Toutefois, une séquence alternative décrite peut être intégrée dans un listage des séquences si le “N” est remplacé par un nucléotide défini de manière spécifique.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

L’intégration des séquences représentant des modes de réalisation qui constituent une partie essentielle de l’invention est fortement conseillée. L’intégration de ces séquences permet d’effectuer une recherche plus approfondie et porte à la connaissance du public l’objet pour lequel la délivrance d’un brevet est demandée.

En ce qui concerne l’exemple ci‑dessus, il est vivement recommandé d’intégrer dans le listage des séquences les trois séquences supplémentaires ci‑après, chacune disposant de son propre numéro d’identification de séquence :

aatgccggag (SEQ ID NO : 12)

aatgccggak (SEQ ID NO : 13)

aatgccggac (SEQ ID NO : 14)

Si les séquences susmentionnées ne sont pas intégrées toutes les trois, le nucléotide qui remplace le “n” devrait être annoté pour décrire les alternatives. Par exemple, si seule la séquence SEQ ID NO : 12 ci‑dessus est intégrée dans le listage des séquences, la clé de caractérisation “misc\_difference” avec l’emplacement de caractéristique “10” doit être employée avec deux qualificateurs du type “replace” lorsque la valeur de l’un serait “g” et la valeur de l’autre serait “c”.

**N.B. :** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**3.k)**, 7.a), 8 et 13

**Exemple 3.k)-3 : Symbole ambigu “n” employé d’une manière non conventionnelle**

Une demande divulgue la séquence : 5’-aatgttggan-3’

où n représente c

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Conformément au paragraphe 3.k), un nucléotide “défini de manière spécifique” est tout nucléotide différent de ceux qui sont représentés par le symbole “n” énumérés dans l’annexe I, section 1, tableau 1.

Dans cet exemple, “n” est employé d’une manière non conventionnelle pour représenter uniquement “c”. La divulgation n’indique pas que “n” est employé d’une manière conventionnelle pour représenter “tout nucléotide”. La séquence doit donc être interprétée comme si le symbole conventionnel équivalent, c’est‑à‑dire “c”, avait été employé dans la séquence (voir l’Introduction au présent document). Il s’ensuit que la séquence énumérée à considérer est la suivante :

5’-aatgttggac-3’

Cette séquence contient 10 nucléotides définis de manière spécifique et doit, conformément au paragraphe 7.a) de la norme ST.26, être intégrée dans le listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence doit être intégrée dans un listage des séquences sous la forme suivante : aatgttggac (SEQ ID NO : 15)

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**3.k)** et 7.a)

**Exemple 3.k)-4 : Les symboles ambigus autres que “n” sont “définis de manière spécifique**

Une demande de brevet décrit la séquence ci‑après :

5’ NNG KNG KNG KAG VCR 3’

où N, K, V et R sont des codes ambigus IUPAC-IUB

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Les codes ambigus IUPAC-IUB correspondent à la liste des symboles de nucléotides définis dans l’annexe I, section 1, tableau 1. Conformément au paragraphe 3.k), un nucléotide “défini de manière spécifique” est tout nucléotide différent de ceux qui sont représentés par le symbole “n” énumérés dans l’annexe I, section 1, tableau 1. “K”, “V” et “R” sont donc des nucléotides “définis de manière spécifique”.

Cette séquence contient 11 nucléotides énumérés et “définis de manière spécifique” et doit, conformément au paragraphe 7.a) de la norme ST.26, être intégrée dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence doit être intégrée dans un listage des séquences sous la forme suivante :

nngkngkngkagvcr (SEQ ID NO : 16)

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~ **3.k)**, 7.a) et 15

**Exemple 3.k)-5 : Abréviation ambiguë “Xaa” employée d’une manière non conventionnelle**

Une demande de brevet décrit la séquence ci‑après :

Xaa-Tyr-Glu-Xaa-Xaa-Xaa-Leu

où Xaa à la position 1 représente tout acide aminé, Xaa à la position 4 représente Lys, Xaa à la position 5 représente Gly et Xaa à la position 6 représente la leucine ou l’isoleucine.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Le peptide énuméré dans la formule contient trois acides aminés définis de manière spécifique aux positions 2, 3 et 7. Le premier acide aminé est représenté par une abréviation conventionnelle, à savoir Xaa, qui représente tout acide aminé. Toutefois, les 4e, 5e et 6e acides aminés sont représentés par une abréviation conventionnelle employée d’une manière non conventionnelle (voir l’Introduction au présent document). De ce fait, l’explication de la séquence donnée dans la divulgation est consultée pour déterminer la définition de “Xaa” à ces positions. Étant donné que les “Xaa” aux positions 4 à 6 désignent un acide aminé particulier, la séquence doit être interprétée comme si les abréviations conventionnelles équivalentes avaient été employées dans la séquence, à savoir Lys, Gly et (Leu or Ile). Il s’ensuit que la séquence contient au moins quatre acides aminés définis de manière spécifique et doit être intégrée dans un listage des séquences comme le prescrit le paragraphe 7.b) de la norme ST.26.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence emploie une abréviation conventionnelle “Xaa” d’une manière non conventionnelle. De ce fait, l’explication de la séquence donnée dans la divulgation est consultée pour déterminer la définition de “Xaa” aux positions 4, 5 et 6. L’explication définit “Xaa” comme une lysine à la position 4, une glycine à la position 5 et une leucine ou isoleucine à la position 6. Les symboles conventionnels pour ces acides aminés sont K, G et J, respectivement. En conséquence, la séquence devrait être représentée dans le listage des séquences sous la forme suivante :

XYEKGJL (SEQ ID NO : 17)

Conformément au paragraphe 27, “X” sera considéré comme l’équivalent de l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire dans le tableau de caractéristiques. Dès lors, si “X” est destiné à représenter “tout acide aminé”, il doit être annoté au moyen de la clé de caractérisation “VARIANT” et d’un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

Chaque fois que possible, chaque “X” devrait être annoté individuellement. Cependant, une région de résidus “X” contigus ou un grand nombre de résidus “X” dispersés dans l’ensemble de la séquence peuvent être décrits conjointement par la clé de caractérisation “VARIANT” en employant la syntaxe “x.y” pour désigner le descripteur d’emplacement, où x et y sont les positions du premier et du dernier résidu “X”, et par un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**3.k)**, 7.b), 26 et 27

### Paragraphe 7.a) – Séquences de nucléotides à intégrer dans un listage

**Exemple 7.a)-1 : Séquence de nucléotides ramifiée**

La description divulgue la séquence de nucléotides ramifiée ci‑après :



où “pnp” est une liaison ou un monomère contenant une fonctionnalité bromoacetylamino;

3’-CA(pnp)CACACA(pnp)CACACA(pnp)CACACACA-(5’)NH—C(=O)CH2 3’ est le segment A;

SP(O-)(=O)CACACAAAAAAAAAAAAAAAAAAAAAAAAA 3’ représente les segments B, C et D; et

SP(O-)(=O)CACATAGGCATCTCCTAGTGCAGGAAGA 3’ est le segment E.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

OUI – les quatre segments verticaux B à E doivent être intégrés dans un listage des séquences.

NON – le segment horizontal A ne doit pas être intégré dans un listage des séquences.

La figure ci‑dessus est un exemple de séquence d’acides nucléiques ramifiée “en peigne” contenant cinq segments linéaires : le segment horizontal A et les quatre segments verticaux B à E.

Conformément au paragraphe 7.a), les ~~portions~~ régions linéaires de séquences de nucléotides ramifiées contenant au moins 10 nucléotides définis de manière spécifique, et dont les nucléotides adjacents sont reliés par une liaison de 3’ à 5’, doivent être intégrées dans un listage des séquences.

Les quatre segments verticaux B à E contiennent chacun plus de 10 nucléotides définis de manière spécifique, et dont les nucléotides adjacents sont reliés par une liaison 3’ à 5’; ils doivent donc tous être intégrés dans un listage des séquences.

S’agissant du segment horizontal A, les ~~portions~~ régions linéaires de la séquence de nucléotides sont reliées par le fragment non nucléotidique “pnp” et chacune de ces portions linéaires reliées contient moins de 10 nucléotides définis de manière spécifique. De ce fait, étant donné qu’aucune ~~portion~~ région du segment A ne contient au moins 10 nucléotides définis de manière spécifique, et dont les nucléotides adjacents soient reliés par une liaison de 3’ à 5’, le paragraphe 7.a) de la norme ST.26 ne prescrit pas leur intégration dans le listage des séquences.

**Question 2 : La norme ST.26 autorise‑t‑elle l’intégration de la ou des séquences?**

**NON**

Conformément au paragraphe 8, “un listage des séquences ne doit contenir aucune séquence comportant moins de 10 nucléotides définis de manière spécifique…”

Aucune ~~portion~~ région du segment A ne contient au moins 10 nucléotides définis de manière spécifique et dont les nucléotides adjacents soient reliés par une liaison de 3’ à 5’; en conséquence, elle ne doit pas être intégrée dans un listage des séquences en tant que séquence distincte disposant de son propre numéro d’identification de séquence.

Toutefois, les segments B, C, D et E peuvent faire l’objet d’une annotation indiquant qu’ils sont reliés au segment A.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Les segments B, C et D sont identiques et doivent être intégrés dans un listage des séquences en tant que séquence unique :

cacacaaaaaaaaaaaaaaaaaaaaaaaaa. (SEQ ID NO : 18)

Le premier “c” de la séquence devrait s’accompagner d’une description supplémentaire en tant que nucléotide modifié à l’aide de la clé de caractérisation “misc\_feature” et du qualificateur de type “note” prenant une valeur qui serait par exemple la suivante : “Cette séquence est l’un des quatre rameaux d’un polynucléotide ramifié.”

Le segment E doit être intégré dans un listage des séquences en tant que séquence unique :

cacataggcatctcctagtgcaggaaga. (SEQ ID NO : 19)

Le premier “c” de la séquence devrait s’accompagner d’une description supplémentaire en tant que nucléotide modifié à l’aide de la clé de caractérisation “misc\_feature” et du qualificateur de type “note” prenant une valeur qui serait par exemple la suivante : “Cette séquence est l’un des quatre rameaux d’un polynucléotide ramifié.”

**Paragraphe(s) pertinent(s) de la ST.26 :** ~~paragraphes~~**7.a)**, 8, 11, 13 et 17

**Exemple 7.a)-2 : Séquence de nucléotides linéaire comportant une structure secondaire**

Une demande de brevet décrit la séquence ci‑après :



où Ψ est la pseudouridine.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La séquence nucléotidique contient 73 nucléotides énumérés et définis de manière spécifique. L’exemple présente donc au moins 10 nucléotides “définis de manière spécifique” et, comme le prescrit le paragraphe 7.a) de la norme ST.26, doit être intégré dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La consultation de la divulgation indique que “Ψ” est l’équivalent de la pseudouridine. Le seul symbole conventionnel qui peut servir à représenter la pseudouridine est “n”; il s’ensuit que le “Ψ” est un symbole non conventionnel employé pour représenter le symbole conventionnel “n” (voir l’Introduction au présent document). En conséquence, la séquence doit être interprétée comme ayant deux symboles “n” à la place des deux symboles “Ψ”.

Le symbole “u” ne doit pas être employé pour représenter l’uracile dans une molécule d’ARN dans le listage des séquences. Conformément au paragraphe 14, le symbole “t” désigne l’uracile dans de l’ARN. La séquence doit être intégrée sous la forme suivante :

gcggatttagctcagctgggagagcgccagactgaatanctggagtcctgtgtncgatccacagaattcgcacca (SEQ ID NO : 20)

La valeur du qualificateur obligatoire “mol\_type” de la clé de caractérisation “source” est “tRNA”. Des informations supplémentaires peuvent être fournies avec la clé de caractérisation “tRNA” et un ou des qualificateurs appropriés.

Les résidus “n” doivent s’accompagner d’une description supplémentaire dans un tableau de caractéristiques en employant la clé de caractérisation “modified\_base” et le qualificateur obligatoire “mod\_base” avec l’abréviation “p” pour la pseudouridine en tant que valeur du qualificateur (voir l’annexe 1, tableau 2).

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**7.a)**, 11, 13, 14, 17, 62, 84 et annexe I, sections 2 et 5, clé de caractérisation 5.43

**Exemple 7.a)-3 : Symboles ambigus des nucléotides employés d’une manière non conventionnelle**

Une demande de brevet décrit la séquence ci‑après :

5’ GATC-MDR-MDR-MDR-MDR-GTAC 3’

L’explication de la séquence donnée dans la divulgation indique par ailleurs ce qui suit : “Un “DR Element” est constitué par la séquence 5’ ATCAGCCAT 3’. Un DR Element mutant, ou MDR, est un élément DR dont les cinq nucléotides du milieu, CAGCC, deviennent après mutation TTTTT.”

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La séquence énumérée emploie le symbole “MDR”. Lorsque l’on ignore si un symbole employé dans une séquence est censé être un symbole conventionnel, c’est‑à‑dire un symbole indiqué dans l’annexe 1, section 3, tableau 3, ou un symbole non conventionnel, l’explication de la séquence donnée dans la divulgation doit être consultée pour le déterminer (voir l’Introduction au présent document). Conformément au tableau 3, “MDR” pourrait être interprété comme désignant trois symboles conventionnels (m = a ou c, d = a ou g ou t/u, r = g ou a) ou une abréviation qui est une notation topologique pour une autre structure.

La consultation de la divulgation indique qu’un élément MDR est l’équivalent de 5’ ATTTTTTAT 3’. Les lettres “MDR” sont considérées comme des symboles conventionnels employés d’une manière non conventionnelle; la séquence doit donc être interprétée comme si elle était divulguée à l’aide des symboles conventionnels équivalents. En conséquence, la séquence énumérée que l’on envisage d’intégrer dans un listage des séquences est la suivante :

5’ GATC ATTTTTTAT ATTTTTTAT ATTTTTTAT ATTTTTTAT GTAC 3’

La séquence énumérée comptant 44 nucléotides définis d’une manière spécifique, elle doit, conformément au paragraphe 7.a) de la norme ST.26, être intégrée dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence doit être intégrée dans un listage des séquences sous la forme suivante :

gatcattttttatattttttatattttttatattttttatgtac (SEQ ID NO : 21)

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**7.a)** et 13

**Exemple 7.a)-4 : Symboles ambigus des nucléotides employés d’une manière non conventionnelle**

Une demande de brevet décrit la séquence ci‑après :

5’ ATTC-N-N-N-N-GTAC 3’

L’explication de la séquence donnée dans la divulgation indique par ailleurs que “N” est constitué par la séquence 5’ ATACGCACT 3’.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La séquence énumérée emploie le symbole “N”. L’explication de la séquence donnée dans la divulgation doit être consultée pour déterminer si le “N” est employé d’une manière conventionnelle ou non conventionnelle (voir l’Introduction au présent document).

La consultation de la divulgation indique que “N” est l’équivalent de 5’ ATACGCACT 3’. Le “N” est donc un symbole conventionnel employé d’une manière non conventionnelle. En conséquence, la séquence doit être interprétée comme si elle était divulguée à l’aide des symboles conventionnels équivalents :

5’ ATTC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3’

La séquence énumérée comptant 44 nucléotides définis d’une manière spécifique, elle doit, conformément au paragraphe 7.a) de la norme ST.26, être intégrée dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence doit être intégrée dans un listage des séquences sous la forme suivante :

attcatacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO : 22)

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**7.a)** et 13

**Exemple 7.a)-5 : Symboles des nucléotides non conventionnels**

Une demande de brevet décrit la séquence ci‑après :

5’ GATC-β-β-β-β-GTAC 3’

L’explication de la séquence donnée dans la divulgation indique par ailleurs que “β” est constitué par la séquence 5’ ATACGCACT 3’.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La séquence énumérée emploie le symbole non conventionnel “β”. L’explication de la séquence donnée dans la divulgation doit être consultée pour déterminer la signification de “β” (voir l’Introduction au présent document).

La consultation de la divulgation indique que “β” est l’équivalent de 5’ ATACGCACT 3’. Le “β” est donc un symbole non conventionnel employé pour représenter une séquence de neuf symboles conventionnels définis d’une manière spécifique. En conséquence, la séquence doit être interprétée comme si elle était divulguée à l’aide des symboles conventionnels équivalents :

5’ GATC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3’

La séquence énumérée comptant 44 nucléotides définis d’une manière spécifique, elle doit, conformément au paragraphe 7.a) de la norme ST.26, être intégrée dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence doit être intégrée dans un listage des séquences sous la forme suivante :

gatcatacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO : 23)

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**7.a)** et 13

**Exemple 7.a)-6 : Symboles de nucléotides non conventionnels**

Une demande de brevet décrit la séquence ci‑après :

5’ GATC-β-β-β-β-GTAC 3’

L’explication de la séquence donnée dans la divulgation indique par ailleurs que “β” est égal à l’adénine, à l’inosine ou à la pseudouridine.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**NON**

La séquence énumérée emploie le symbole non conventionnel “β”. L’explication de la séquence donnée dans la divulgation doit être consultée pour déterminer la signification de “β” (voir l’Introduction au présent document).

La consultation de la divulgation indique que “β” est l’équivalent de l’adénine, de l’inosine ou de la pseudouridine. Le seul symbole conventionnel qui peut être employé pour représenter “adénine, inosine ou pseudouridine” est “n”; le “β” est donc un symbole non conventionnel employé pour représenter le symbole conventionnel “n”. En conséquence, la séquence doit être interprétée comme ayant quatre symboles “n” à la place des quatre symboles “β” :

5’ GATC-N-N-N-N-GTAC 3’

La séquence énumérée ne comportant que huit nucléotides définis de manière spécifique, elle ne doit pas, conformément au paragraphe 7.a) de la norme ST.26, être intégrée dans un listage des séquences.

**Question 2 : La norme ST.26 autorise‑t‑elle l’intégration de la ou des séquences?**

**NON**

La séquence énumérée, 5’ GATC-N-N-N-N-GTAC 3’, ne doit pas être intégrée dans un listage des séquences.

Toutefois, une séquence alternative divulguée peut être intégrée dans un listage des séquences si au moins deux des symboles “n” sont remplacés par l’adénine, ce qui donnerait une séquence d’au moins 10 nucléotides définis de manière spécifique.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Une représentation autorisée possible est la suivante :

gatcaaaagtac (SEQ ID NO : 24)

Dans l’exemple ci‑dessus, les quatre nucléotides d’adénine qui remplacent les symboles β devraient s’accompagner d’une annotation indiquant que ces positions pourraient être remplacées par l’inosine ou la pseudouridine.

La clé de caractérisation “misc\_difference” devrait être employée avec un emplacement de caractéristique 5‑8 et un qualificateur du type “note” prenant la valeur “Un nucléotide aux positions 5‑8 peut être remplacé par l’inosine ou la pseudouridine”, par exemple. Étant donné que ces alternatives sont des nucléotides modifiés, la clé de caractérisation “modified\_base” et le qualificateur “mod\_base” seraient requis. La valeur du qualificateur “mod\_base” peut être “OTHER” avec un qualificateur du type “note” et la valeur de “i ou p”.

D’autres permutations sont possibles.

**N.B. :** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**7.a)**, 8, 13 et 17

### Paragraphe 7.b) – Séquences d’acides aminés à intégrer dans un listage

**Exemple 7.b)-1 : Au moins quatre acides aminés définis de manière spécifique**

XXXXXXXXDXXXXXXXXXXFXXXXXXXXXXXXXXXXXXXXXXXXXXXXAXXXXXXXXXXXXXXXXXXXGXXXXX

où X = tout acide aminé

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Le peptide énuméré contient quatre acides aminés définis de manière spécifique. Le symbole “X” est employé de manière conventionnelle pour représenter les autres acides aminés comme étant tout acide aminé (voir l’Introduction au présent document).

Étant donné qu’il y a quatre acides aminés définis de manière spécifique, à savoir Asp, Phe, Ala et Gly, le paragraphe 7.b) de la norme ST.26 prescrit l’intégration de la séquence dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence doit être représentée sous la forme suivante :

XXXXXXXXDXXXXXXXXXXFXXXXXXXXXXXXXXXXXXXXXXXXXXXXAXXXXXXXXXXXXXXXXXXXGXXXXX (SEQ ID NO : 25)

~~Since “X” can be any amino acid, annotation of the “X” residues is not required under paragraph 27.~~ Conformément au paragraphe 27, “X” sera considéré comme l’équivalent de l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire dans le tableau de caractéristiques. Dès lors, si “X” est destiné à représenter “tout acide aminé”, il doit être annoté au moyen de la clé de caractérisation “VARIANT” et d’un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

Chaque fois que possible, chaque “X” devrait être annoté individuellement. Cependant, une région de résidus “X” contigus ou un grand nombre de résidus “X” dispersés dans l’ensemble de la séquence peuvent être décrits conjointement par la clé de caractérisation “VARIANT” en employant la syntaxe “x.y” pour désigner le descripteur d’emplacement, où x et y sont les positions du premier et du dernier résidu “X”, et par un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**7.b)**, 8 et 27

**Exemple 7.b)-2 : Séquence d’acides aminés ramifiée**

La demande décrit une séquence ramifiée dans laquelle les résidus de lysine sont utilisés comme squelette pour former huit rameaux auxquels sont reliées plusieurs chaînes peptidiques linéaires. La lysine est un acide aminé dibasique, qui dispose de deux sites de liaison peptidique. Le peptide est illustré comme suit :

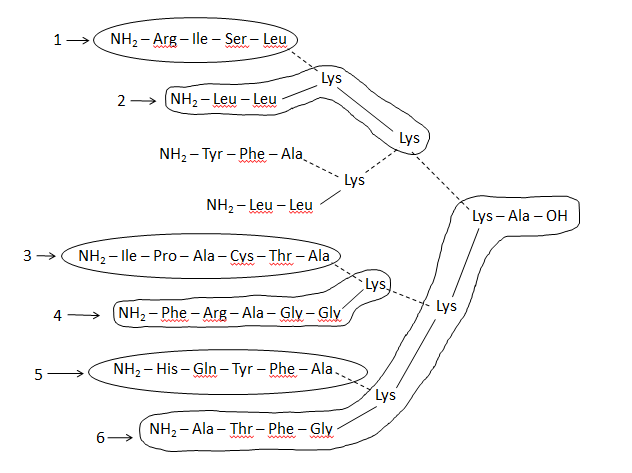


Dans le peptide ramifié ci‑dessus, les liens entre la lysine et un autre acide aminé décrits par représentent une liaison amide entre la terminaison amine de la lysine ~~L~~ et la terminaison carboxyle de l’acide aminé lié. Les liens décrits par représentent une liaison amide entre la chaîne latérale de la lysine ~~L~~ et la terminaison carboxyle de l’acide aminé lié.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

L’exemple divulgue une séquence ramifiée où les résidus de lysine sont utilisés comme un squelette. Le paragraphe 7.b) prescrit l’intégration dans un listage des séquences de la ~~portion~~ région non ramifiée ou linéaire de la séquence, qui contient au moins quatre acides aminés définis de manière spécifique. Dans l’exemple ci‑dessus, les ~~portions~~ régions linéaires du peptide ramifié qui comportent au moins quatre acides aminés définis de manière spécifique sont cerclées :



Le paragraphe 7.b) de la norme ST.26 prescrit l’intégration des peptides 1 à 6 ci‑dessus dans un listage des séquences.

Les peptides dont l’intégration dans un listage des séquences n’est pas prescrite~~, et de fait est même interdite,~~ sont les suivants :

YFA

LLK

**Question 2 : La norme ST.26 autorise‑t‑elle l’intégration de la ou des séquences?**

**NON**

Conformément au paragraphe 8, “un listage des séquences ne doit contenir aucune séquence comportant moins de quatre acides aminés définis de manière spécifique

Les peptides YFA et LLK ne contiennent chacun que trois acides aminés définis de manière spécifique et ne doivent donc pas être inclus dans un listage des séquences en tant que séquences distinctes disposant de leur propre numéro d’identification de séquence.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Les peptides 1 à 6 doivent être représentés par des identificateurs de séquence distincts :

RISL (SEQ ID NO : 26)

LLKK (SEQ ID NO : 27)

IPACTA (SEQ ID NO : 28)

FRAGGK (SEQ ID NO : 29)

HQYFA (SEQ ID NO : 30)

ATFGKKKA (SEQ ID NO : 31)

La liaison croisée doit de préférence être indiquée à l’aide de la clé de caractérisation “SITE” et du qualificateur obligatoire “NOTE”, qui peut par exemple prendre la valeur “This sequence is one part of a branched ~~peptide~~amino acid sequence” (“Cette séquence fait partie d’une séquence ~~de peptides~~d’acides aminés ramifiée”). Selon le paragraphe 29 de la norme ST.26, les numéros d’identification de séquence 27, 29 et 31 doivent comporter pour chaque lysine une annotation indiquant qu’il s’agit d’un acide aminé modifié; on utilisera la clé de caractérisation “SITE” ainsi que le qualificateur “NOTE” pour indiquer que la chaîne latérale de la lysine est reliée à une autre séquence par une liaison amide. Les numéros d’identification de séquence 26, 28 et 30 devraient de préférence s’accompagner d’une annotation indiquant que le terminal C d’un acide aminé est lié à une autre séquence; on utilisera à cette fin la clé de caractérisation “SITE” ainsi que le qualificateur “NOTE”.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**7.b)**, 8 à 26, 29, 30 et 31

**Exemple 7.b)-3 : Séquence d’acides aminés ramifiée**

Un peptide de la séquence ci‑après :



La liaison entre le résidu terminal de glycine de la séquence inférieure et la lysine de la séquence supérieure s’établit par l’intermédiaire d’une liaison amide entre la terminaison carboxyle de la glycine et la chaîne latérale du terminal amino de la lysine.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La ~~portion~~ région non ramifiée ou linéaire d’une séquence contenant au moins quatre acides aminés définis de manière spécifique doit être intégrée dans un listage des séquences. Dans l’exemple ci‑dessus, les ~~portions~~ régions linéaires du peptide ramifié qui comptent plus de quatre acides aminés sont les suivantes :



Le paragraphe 7.b) de la norme ST.26 prescrit l’intégration des ~~peptides~~ séquences 1 et 2 dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Les ~~peptides~~ séquences 1 et 2 doivent être représentés par des identificateurs de séquence distincts :

DGSAKKKK (SEQ ID NO : 32)

AASHG (SEQ ID NO : 33)

La séquence DGSAKKKK ~~devrait de preference~~doit s’accompagner d’une annotation indiquant que la ~~5~~~~e~~ lysine en position numéro 5 est un acide aminé modifié; on utilisera la clé de caractérisation “SITE” ainsi que le qualificateur “NOTE” pour indiquer que la chaîne latérale de la lysine est liée ~~au peptide AASHG~~ à une autre séquence par une liaison amide. La séquence AASHG devrait de préférence s’accompagner d’une annotation indiquant que la ~~5e~~ glycine en position numéro 5 est liée à ~~la séquence DGSAKKK~~ une autre séquence en utilisant la clé de caractérisation “SITE” ainsi que le qualificateur “NOTE”.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**7.b)**, 26, 29, 30 et 31

**Exemple 7.b)-4 : Peptide cyclique contenant une séquence d’acides aminés ramifiée**

La structure suivante est divulguée dans une demande de brevet :



La cystéine et la leucine de la structure cyclique sont liées par la chaîne latérale de la Cys et la terminaison carbonyle de la Leu.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

La structure ci‑dessous est une séquence d’acides aminés cyclique et ramifiée contenant les acides aminés suivants :



Comme la chaîne latérale de la Cys et la terminaison carbonyle de la Leu participent à la cyclisation, la terminaison N du peptide cyclique se trouve au Cys-1.

**OUI – partie cyclique du peptide**

La norme ST.26 prévoit au paragraphe 7.b) que la région linéaire d’une séquence ramifiée contenant au moins quatre acides aminés définis de manière spécifique, et dont les acides aminés forment un seul squelette de peptides, doit figurer dans un listage des séquences. Dans l’exemple ci‑dessus, la partie cyclique du peptide ramifié compte plus de quatre acides aminés; elle doit donc être intégrée dans un listage des séquences.

**NON – ramification tripeptide du peptide**

Il n’est pas obligatoire d’intégrer la ramification du tripeptide Ala-Leu-Glu dans le listage des séquences.

**Question 2 : La norme ST.26 autorise‑t‑elle l’intégration de la ou des séquences?**

**NON**

Conformément au paragraphe 8, “un listage des séquences ne doit contenir aucune séquence comportant moins de quatre acides aminés définis de manière spécifique

La ramification du tripeptide ne contenant que trois acides aminés définis de manière spécifique, elle ne doit pas être intégrée dans le listage des séquences en tant que séquence distincte disposant de son propre numéro d’identification.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Bien que cet exemple illustre un peptide de configuration circulaire, l’anneau ne se compose pas uniquement de résidus d’acides aminés liés par des peptides, conformément au paragraphe 25. Comme la cyclisation de la séquence d’acides aminés s’effectue au moyen de la chaîne latérale de la cystéine (Cys) et de la terminaison carboxyle de la ~~L~~leucine (Leu), il faut attribuer le numéro de position 1 à la cystéine dans la partie cyclique du peptide. La séquence doit donc être représentée comme suit :

CALRDKL (SEQ ID NO : 89).

Comme l’indique la figure ci‑dessus, la séquence d’acides aminés est cyclisée par une conjugaison thioester entre la chaîne latérale de la cystéine et la terminaison carboxyle de la leucine. Il faut employer la clé de caractérisation “SITE” pour décrire la cystéine modifiée, qui forme la liaison intrachaîne avec la leucine. Il faut appliquer l’opérateur d’emplacement “join” aux descripteurs d’emplacement pour désigner les résidus intervenant dans la liaison, par exemple “join(1,7)”. Le qualificateur obligatoire “NOTE” doit indiquer la nature de la liaison, par exemple “~~C~~cysteine ~~L~~leucine thioester (Cys-Leu)” pour préciser que le Cys-1 et le Leu-7 sont reliés par une liaison thioester. De plus, la lysine en position numéro 6 doit être annotée pour indiquer qu’elle a été modifiée; on emploie à cette fin la clé de caractérisation “SITE” ainsi que le qualificateur obligatoire “NOTE”, la valeur de ce dernier indiquant que le tripeptide ALE est relié à la chaîne latérale de la lysine.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~**7.b)**, 8, 25, 26, 29, 30, 31, 67 et 68

**Exemple 7.b)-5 : Peptide cyclique contenant une séquence d’acides aminés ramifiée**

Le peptide cyclique ramifié suivant est divulgué dans une demande de brevet :



Le Ser et le Lys sont reliés par une liaison amide entre la terminaison carboxy de la sérine et l’amine dans la chaîne latérale du Lys.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Le paragraphe 7.b) prévoit que toute séquence contenant au moins quatre acides aminés définis de manière spécifique et pouvant être représentée comme une région linéaire dans une séquence ramifiée doit figurer dans le listage des séquences. Dans l’exemple ci‑dessus, le peptide contient une région cyclique dans laquelle les acides aminés sont reliés par des liaisons peptidiques, ainsi qu’une région ramifiée qui est liée à une chaîne latérale du Lys dans la région cyclique. Les régions de ce peptide ramifié qui peuvent être représentées comme des régions linéaires et qui contiennent au moins quatre acides aminés définis de manière spécifique sont les suivantes :



La norme ST.26 prescrit d’intégrer les séquences 1 et 2 de ce peptide ramifié cyclique dans un listage des séquences, chacune de ces séquences devant disposer de son propre numéro d’identification.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence 1 doit être représentée ainsi :

LRDQS (SEQ. ID. NO : 90)

La séquence devrait de préférence être annotée au moyen de la clé de caractérisation “SITE” et du qualificateur “NOTE” pour indiquer que la sérine en position 5 est reliée à une autre séquence par une liaison amide entre le Ser et une chaîne latérale du Lys dans l’autre séquence.

La séquence 2 est un peptide cyclique. Selon le paragraphe 25, lorsqu’une séquence d’acides aminés a une configuration cyclique et n’a pas de terminaison amine ou carboxy, le requérant doit choisir le résidu d’acides aminés qu’il entend placer en position numéro 1. La séquence peut alors être représentée ainsi :

ALFKNG (SEQ. ID. NO : 91)

Tout autre acide aminé de la séquence peut également être placé en position de résidu numéro 1. La séquence ALFKNG doit être décrite plus en détail au moyen de la clé de caractérisation “SITE” et du qualificateur “NOTE” pour indiquer que la chaîne latérale du Lys en position de résidu numéro 4 est reliée par une liaison amide à une autre séquence. Cette liaison de la chaîne latérale modifie le Lys, et selon le paragraphe 30 de la norme ST.26, tout acide aminé modifié doit être accompagné d’une description supplémentaire dans le tableau de caractéristiques. De plus, une clé de caractérisation “REGION” et un qualificateur “NOTE” devraient être prévus pour montrer que le peptide ALFKNG est circulaire.

**Paragraphe(s) pertinent(s) de la norme  ST.26 :** ~~paragraphes~~**7.b)**, 25, 26, 30 et 31

### Paragraphe 11.a) – Séquence de nucléotides représentée par deux brins de codage – entièrement complémentaires

**Exemple 11.a)-1 : Séquence de nucléotides représentée par deux brins de codage – mêmes longueurs**

Une demande de brevet décrit la séquence d’ADN représentée par deux brins de codage ci‑après :

3’-CCGGTTAACGCTA-5’

5’-GGCCAATTGCGAT-3’

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Chaque séquence nucléotidique énumérée comporte plus de 10 nucléotides définis de manière spécifique. Un brin de codage au moins doit être intégré dans le listage des séquences, car les deux brins de cette séquence nucléotidique représentée par deux brins de codage sont entièrement complémentaires l’un de l’autre.

**Question 2 : La norme ST.26 autorise‑t‑elle l’intégration de la ou des séquences?**

**OUI**

Si la séquence d’un seul brin de codage doit être intégrée dans le listage des séquences, les séquences des deux brins de codage peuvent y être intégrées, chacune disposant de son propre numéro d’identification de séquence.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence d’ADN représentée par deux brins de codage doit être représentée sous la forme d’une seule séquence ou de deux séquences distinctes. Chaque séquence intégrée dans le listage des séquences doit être représentée dans le sens 5’-3’ et disposer de son propre numéro d’identification de séquence.

atcgcaattggcc (brin supérieur) (SEQ ID NO : 34)

et/ou

ggccaattgcgat (brin inférieur) (SEQ ID NO : 35)

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7.a), **11.a)** et 13

### Paragraphe 11.b) – Séquence de nucléotides représentée par deux brins de codage – non entièrement complémentaires

**Exemple 11.b)-1 : Séquence de nucléotides représentée par deux brins de codage – différentes longueurs**

Une demande de brevet contient le dessin et la légende ci‑après :

5’-tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc-3’

||||||||||||||||

gggtaactgantccgc

Le promoteur du gène humain ABC1 (brin supérieur) lié par une sonde d’ANP (brin supérieur). Lorsque “n” dans la sonde d’ANP est une base d’ANP universelle choisie parmi le groupe comprenant le 5‑nitroindole et le 3‑nitroindole.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI** – le promoteur de l’ABC1 (brin supérieur)

Contenant plus de 10 nucléotides énumérés et “définis de manière spécifique”, le brin supérieur doit être intégré dans un listage des séquences.

**OUI** – la sonde ANP (brin inférieur)

Le brin inférieur doit également être intégré dans le listage des séquences en disposant de son propre numéro d’identification de séquence, car les deux brins ne sont pas entièrement complémentaires l’un de l’autre. Les résidus individuels qui constituent un ANP ou “acide nucléique peptidique” sont considérés comme des nucléotides conformément au paragraphe 3.g) de la norme ST.26. Il s’ensuit que le brin inférieur contient plus de 10 nucléotides énumérés et “définis de manière spécifique”; de ce fait, il doit être intégré dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Le brin supérieur doit être intégré dans un listage des séquences sous la forme suivante :

tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc (SEQ ID NO : 36)

Le brin supérieur est un acide nucléique peptidique et, de ce fait, il ne possède pas d’extrémité 3’ ni d’extrémité 5’. Conformément au paragraphe 11, il doit être intégré dans un listage des séquences “de gauche à droite de manière à reproduire le sens 5’-3’.” Il doit donc être intégré dans un listage des séquences sous la forme suivante :

cgcctnagtcaatggg (SEQ ID NO : 37)

Le qualificateur du type “organism” de la clé de caractérisation “source” doit prendre la valeur “synthetic construct” et le qualificateur obligatoire du type “mol\_type” la valeur “other DNA”. Le brin inférieur doit être décrit dans un tableau de caractéristiques à l’aide de la clé de caractérisation “modified\_base” et du qualificateur obligatoire “mod\_base” avec l’abréviation “OTHER”. Un qualificateur du type “note” doit être intégré avec le nom complet non abrégé des nucléotides modifiés, tels que les “N-(2-aminoethyl) glycine nucleosides”.

Le résidu “n” doit s’accompagner d’une description supplémentaire dans un tableau de caractéristiques à l’aide de la clé de caractérisation “modified\_base” et du qualificateur obligatoire “mod\_base” avec l’abréviation “OTHER”. Un qualificateur du type “note” doit être intégré avec le nom complet non abrégé du nucléotide modifié : “N-(2-aminoethyl) glycine 5-nitroindole ou N-(2-aminoethyl) glycine 3-nitroindole”.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~3.g), 7.a), **11.b)**, 17 et 18

**Exemple 11.b)-2 : Séquence de nucléotides représentée par deux brins de codage – absence de segment d’appariement de bases**

Une demande de brevet décrit la séquence d’ADN représentée par deux brins de codage ci‑après :

3’-CCGGTTAGCTTATACGCTAGGGCTA-5’

||||||| ||||||||||||

5’-GGCCAATATGGCTTGCGATCCCGAT-3’

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Chaque brin de la séquence nucléotidique énumérée représentée par deux brins de codage contient plus de 10 nucléotides définis de manière spécifique. Les deux brins doivent être intégrés dans le listage des séquences, chacun disposant de son propre numéro d’identification de séquence, car les deux brins ne sont pas entièrement complémentaires l’un de l’autre.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence de chaque brin doit être représentée dans le sens 5’-3’ et disposer de son propre numéro d’identification :

atcgggatcgcatattcgattggcc (brin supérieur) (SEQ ID NO : 38)

and

ggccaatatggcttgcgatcccgat (brin inférieur) (SEQ ID NO : 39)

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7.a), **11.b)** et 13

### Paragraphe 14 – Le symbole “t”désigne l’uracile dans de l’ARN

**Exemple 14-1 : Le symbole “t” représente l’uracile dans de l’ARN**

Une demande de brevet décrit le composé ci‑après :

segment A: ccugucgt-3’

segment B: uaguuguagaggccugucct-5’



où le segment A et le segment B sont des séquences d’ARN.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI – segment B**

**NON – segment A**

La séquence énumérée contient deux segments de nucléotides définis de manière spécifique séparés par le “lieur” suivant :



Le lieur n’est pas un nucléotide au sens du paragraphe 3.g); de ce fait, chaque segment doit être considéré comme une séquence distincte. Le segment B contenant plus de 10 nucléotides définis de manière spécifique, le paragraphe 7.a) de la norme ST.26 en prescrit l’intégration dans un listage des séquences. Le segment A ne contenant que huit nucléotides définis de manière spécifique, la norme n’en prescrit pas l’intégration dans un listage des séquences.

**Question 2 : La norme ST.26 autorise‑t‑elle l’intégration de la ou des séquences?**

**NON**

Le segment A contenant moins de 10 nucléotides définis de manière spécifique et, selon le paragraphe 8 de la norme ST.26, ~~donc~~ il ne doit pas être intégré dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Le segment B est une molécule d’ARN; de ce fait, l’élément “INSDSeq\_moltype” doit être “ARN.” Le symbole “u” ne doit pas être utilisé pour représenter l’uracile dans une molécule d’ARN dans un listage des séquences. Selon le paragraphe 14, le symbole “t” désigne l’uracile dans de l’ARN. Le segment B doit donc être intégré dans le listage des séquences de la manière suivante :

tcctgtccggagatgttgat (SEQ ID NO : 40)

La thymine dans de l’ARN est considérée comme un nucléotide modifié, c’est‑à‑dire de l’uracile modifiée, et doit être représentée dans la séquence sous la forme “t” et s’accompagner d’une description supplémentaire dans un tableau de caractéristiques. En conséquence, la thymine à la position 1 doit s’accompagner d’une description supplémentaire à l’aide de la clé de caractérisation “modified\_base”, du qualificateur du type “mod\_base” prenant la valeur “OTHER” et d’un qualificateur du type “note” prenant la valeur “thymine”.

La thymine, c’est‑à‑dire l’uracile modifiée, à la position 1 devrait également s’accompagner d’une description supplémentaire dans un tableau de caractéristiques à l’aide de la clé de caractérisation “misc\_feature” et d’un qualificateur du type “note” prenant, par exemple, la valeur “~~ccugucgt (segment A) se fixe à son extrémité 3’ à un lieur qui se fixe à~~l’oxygène 5’ de la thymidine est fixé à une autre séquence de nucléotides par ~~Le~~le lieur, qui est l’acide (4-(3-hydroxybenzamido)butyl) phosphinique. Chaque fois que possible, le qualificateur “note” peut directement prendre la valeur de l’autre séquence.

**Paragraphe(s) pertinent(s) de la ST.26 :** ~~paragraphes~~3.g), 7.a), 8, 13, **14**, 19 et 54

### Paragraphe 27 – Il faut choisir le symbole ambigu le plus le plus restrictif

**Exemple 27-1 : Formule topologique pour une séquence de nucléotides**

(GGGz)2

où z représente tout acide aminé.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La séquence est divulguée sous la forme d’une formule. (GGGz)2 ne fait que représenter la séquence GGGzGGGz d’une manière abrégée. Par convention, on commence par amplifier une séquence et on détermine ensuite la définition de toute variable, à savoir “z”.

La séquence utilise le symbole non conventionnel “z”. La définition de “z” doit être déterminée à partir de l’explication de la séquence donnée dans la divulgation, qui définit ce symbole comme tout acide aminé (voir l’Introduction au présent document). L’exemple ne contient aucune contrainte pour “z”, comme celle de devoir être identique à chaque occurrence.

De ce fait, “z” est l’équivalent du symbole conventionnel “X”, et le peptide de l’exemple contient huit acides aminés énumérés, dont six sont des résidus de glycine définis de manière spécifique. Le paragraphe 7.b) de la norme ST.26 prescrit l’intégration de la séquence dans un listage des séquences sous la forme d’une séquence unique disposant d’un numéro d’identification de séquence unique.

On notera que la séquence reste visée par le paragraphe 7.b) en dépit du fait que les résidus énumérés et définis de manière spécifique ne sont pas contigus.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence utilise le symbole non conventionnel “z”, qui, aux termes de la divulgation, représente tout acide aminé. Le symbole conventionnel utilisé pour représenter “tout acide aminé” est “X”. Il s’ensuit que la séquence doit être représentée sous la forme de la séquence amplifiée unique suivante :

GGGXGGGX (SEQ ID NO : 41)

Conformément au paragraphe 27, “X” sera considéré comme l’équivalent de l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire dans le tableau de caractéristiques. Dès lors, si “X” est destiné à représenter “tout acide aminé”, il doit être annoté au moyen de la clé de caractérisation “VARIANT” et d’un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

Chaque fois que possible, chaque “X” devrait être annoté individuellement. Cependant, une région de résidus “X” contigus ou un grand nombre de résidus “X” dispersés dans l’ensemble de la séquence peuvent être décrits conjointement par la clé de caractérisation “VARIANT” en employant la syntaxe “x..y” pour désigner le descripteur d’emplacement, où x et y sont les positions du premier et du dernier résidu “X”, et par un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

D’autre part, l’exemple ne divulgue pas que “z” est le même acide aminé aux deux positions dans la séquence amplifiée. Toutefois, si “z” est divulgué comme le même acide aminé aux deux positions, il faudrait utiliser une clé de caractérisation “VARIANT” et un qualificateur du type “NOTE” indiquant que “X” aux positions 4 et 8 peut représenter tous acides aminés dès l’instant qu’ils sont identiques aux deux positions.

Paragraphe(s) pertinent(s) de la norme : ~~paragraphes~~3.c), 7.b), 26 et 27

**Exemple 27-2 : Formule topologique – moins de quatre acides aminés définis de manière spécifique**

Un peptide de la formule (Gly-Gly-Gly-z)n

La divulgation indique également que z représente tout acide aminé et que

i) la variable n représente toute longueur; ou que

ii) la variable n est 2-100, de préférence 3

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**NON**

L’examen des deux modes de réalisation divulgués i) et ii) du peptide énuméré de la formule montre que “n” peut représenter “toute longueur”; de ce fait, le mode de réalisation le plus englobant de “n” est indéterminé. Étant donné que “n” est indéterminé, le peptide de la formule ne peut pas être amplifié de manière à avoir une longueur précise, si bien qu’il faut prendre en compte la formule non amplifiée.

Le peptide énuméré dans la formule non amplifiée (“n” = 1) contient trois acides aminés définis de manière spécifique, dont chacun est Gly, et le symbole “z”. Par convention, “Z” est le symbole représentant la “glutamine ou (l’)acide glutamique”; toutefois, l’exemple définit “z” comme “tout acide aminé” (voir l’Introduction au présent document). En vertu de la norme ST.26, un acide aminé qui n’est pas défini de manière spécifique est représenté par “X”. Il découle de cette analyse que le peptide énuméré, c’est‑à‑dire GGGX, ne contient pas quatre résidus d’acides aminés définis de manière spécifique. En conséquence, le paragraphe 7.b) de la norme ST.26 ne prescrit pas l’intégration, en dépit du fait que “n” est également défini comme représentant des valeurs numériques particulières dans certains modes de réalisation.

**Question 2 : La norme ST.26 autorise‑t‑elle l’intégration de la ou des séquences?**

**OUI**

L’exemple contient une valeur numérique particulière pour la variable “n”, à savoir une limite inférieure de 2, une limite supérieure de 100 et une valeur exacte de 3. Toute séquence contenant au moins quatre acides aminés définis de manière spécifique peut être intégrée dans le listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Une séquence contenant 100 copies de GGGX est préférée (SEQ ID NO : 42). Il faudrait prévoir une annotation supplémentaire indiquant que l’on pourrait supprimer jusqu’à 98 copies de GGGX. L’intégration de modes de réalisation particuliers supplémentaires qui constituent une partie essentielle de l’invention est fortement conseillée.

Conformément au paragraphe 27, “X” sera considéré comme l’équivalent de l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire dans le tableau de caractéristiques. Dès lors, si “X” est destiné à représenter “tout acide aminé”, il doit être annoté au moyen de la clé de caractérisation “VARIANT” et d’un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

Chaque fois que possible, chaque “X” devrait être annoté individuellement. Cependant, une région de résidus “X” contigus ou un grand nombre de résidus “X” dispersés dans l’ensemble de la séquence peuvent être décrits conjointement par la clé de caractérisation “VARIANT” en employant la syntaxe “x.y” pour désigner le descripteur d’emplacement, où x et y sont les positions du premier et du dernier résidu “X”, et par un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

N.B. : La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 : ~~paragraphes~~**3.c), 7.b) et **27**

**Exemple 27-3 : Formule topologique – au moins quatre acides aminés définis d’une manière spécifique**

Un peptide de la formule (Gly-Gly-Gly-z)n

où z est tout acide aminé et la variable n est compris entre 2 et 100, et est de préférence 3.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Le peptide énuméré dans la formule contient trois acides aminés définis de manière spécifique, dont chacun est Gly, et le symbole “z”. Par convention, “Z” est le symbole représentant la “glutamine ou (l’)acide glutamique”; toutefois, la description, dans cet exemple, définit “z” comme “tout acide aminé” (voir l’Introduction au présent document). En vertu de la norme ST.26, un acide aminé qui n’est pas défini de manière spécifique est représenté par “X”. Il découle de cette analyse que le peptide répété énuméré ne contient pas quatre acides aminés définis de manière spécifique. Toutefois, la description indique une valeur numérique particulière pour la variable “n”, c’est‑à‑dire une limite inférieure de 2 et une limite supérieure de 100. En conséquence, l’exemple divulgue un peptide contenant aux moins six acides aminés définis de manière spécifique dans la séquence GGGzGGGz, qui doit, conformément à la norme ST.26, être intégrée dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Étant donné que “z” représente tout acide aminé, le symbole conventionnel utilisé pour représenter le quatrième et le huitième acide aminé est “X.”

La norme ST.26 prescrit l’intégration dans un listage des séquences de la seule séquence qui a été énumérée par ses résidus. De ce fait, au moins une séquence contenant 2, 3 ou 100 copies de GGGX doit être intégrée dans le listage des séquences; toutefois, la séquence la plus englobante qui contient 100 copies de GGGX est préférée (SEQ ID NO : 42) (voir l’Introduction au présent document). Dans ce dernier cas, une annotation supplémentaire pourrait indiquer que 98 copies de GGGX au maximum pourraient être supprimées. Il est fortement conseillé d’intégrer deux séquences supplémentaires contenant 2 et 3 copies de GGGX, respectivement (SEQ ID NO : 44‑45).

Conformément au paragraphe 27, “X” sera considéré comme l’équivalent de l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire dans le tableau de caractéristiques. Dès lors, si “X” est destiné à représenter “tout acide aminé”, il doit être annoté au moyen de la clé de caractérisation “VARIANT” et d’un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

Chaque fois que possible, chaque “X” devrait être annoté individuellement. Cependant, une région de résidus “X” contigus ou un grand nombre de résidus “X” dispersés dans l’ensemble de la séquence peuvent être décrits conjointement par la clé de caractérisation “VARIANT” en employant la syntaxe “x..y” pour désigner le descripteur d’emplacement, où x et y sont les positions du premier et du dernier résidu “X”, et par un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

D’autre part, l’exemple ne divulgue pas que la variable “z” est la même dans chacune des deux occurrences dans la séquence amplifiée. Toutefois, si “z” est divulgué comme étant le même acide aminé dans tous les emplacements, une clé de caractérisation VARIANT et un qualificateur du type NOTE devraient indiquer que “X” à toutes les positions peut représenter tous acides aminés, dès l’instant qu’ils sont les mêmes à tous les emplacements.

**N.B.** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~3.c), 7.b), 26, et **27**

### Paragraphe 28 – Séquences d’acides aminés séparées par des symboles internes de fin

**Exemple 28-1 : Codage de la séquence de nucléotides et séquence d’acides aminés codée**

Une demande de brevet décrit les séquences ci‑après :



**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La demande décrit une séquence nucléotidique contenant des codons de terminaison et codant trois séquences d’acides aminés distinctes.

Contenant plus de 10 nucléotides définis de manière spécifique, la séquence nucléotidique énumérée doit être intégrée dans un listage des séquences comme une séquence unique.

En ce qui concerne les séquences d’acides aminés codées, le paragraphe 28 prescrit l’intégration comme séquences distinctes des séquences d’acides aminés séparées par un symbole interne de fin tel qu’un espace blanc. Étant donné que chaque “Protéine A”, “Protéine B” et “Protéine C” contient au moins quatre acides aminés définis de manière spécifique, elles doivent chacune, conformément au paragraphe 7.b) de la norme ST.26, être intégrées dans un listage des séquences et disposer de leur propre numéro d’identification de séquence.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence nucléotidique doit être intégrée dans un listage des séquences sous la forme suivante :

caattcagggtggtgaatatggcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatggaaagcgggcagtgaatgaccatgattacggattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacattggtgtcaaaaataataataaccggatgtactatttatccctgatgctgcgtcgtcaggtgaatgaagtcgcttaagcaatcaatgtcggatgcggcgcgacgcttatccgaccaacatatcataa. (SEQ ID NO : 46)

La séquence nucléotidique devrait s’accompagner d’une description supplémentaire utilisant une clé de caractérisation “CDS” pour chacune des trois protéines, et l’élément INSDFeature\_location ~~devrait~~doit indiquer l’emplacement de chaque séquence de codage, y compris le codon d’arrêt. De plus, il faudrait employer, en association avec chaque clé de caractérisation “CDS”, le qualificateur “translation” qui prenne comme valeur la séquence d’acides aminés de la protéine. La demande ne divulgue pas le tableau du code génétique qui est appliqué à la traduction (voir annexe 1, section 9, tableau 5). Si le tableau de codes normalisés est appliqué, le qualificateur “transl\_table” n’est pas nécessaire; toutefois, si un tableau du code génétique différent est appliqué, il faut indiquer la valeur appropriée figurant dans le tableau 5 pour le qualificateur “transl\_table”. Enfin, il faut intégrer le qualificateur “protein \_id” dont la valeur indiquera le numéro d’identification de séquence de chacune des séquences d’acides aminés traduites.

Les séquences d’acides aminés doivent être intégrées comme séquences distinctes disposant chacune de leur propre numéro d’identification de séquence :

MAPNTQTASPRALADSLMQLARQVSRLESGQ (SEQ ID NO : 47)

MTMITDSLAVVLQRRDWENPGVTQLNRLAAHWCQK (SEQ ID NO : 48)

MLRRQVNEVA (SEQ ID NO : 49)

NOTE : Voir “Exemple 90‑1 Séquence d’acides aminés codée selon une séquence de codage avec introns” pour un exemple de séquence d’acides aminés traduite représentée en tant que séquence unique.

P**aragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7, 26, **28**, 57, 87 à 90

### Paragraphe 29 – Représentation d’un acide aminé modifié “other”

**Exemple 29-1 : Symbole ambigu le plus restrictif pour un acide aminé “other”**

Une demande de brevet décrit la séquence ci‑après :

Ala-Hse-X1-X2-X3-X4-Tyr-Leu-Gly-Ser

où, X1= Ala ou Gly,

X2= Ala ou Gly,

X3= Ala ou Gly,

X4= Ala ou Gly, et

Hse = homosérine

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Le peptide énuméré contient cinq acides aminés définis de manière spécifique. Le symbole “X” est employé par convention pour représenter deux acides aminés alternatifs (voir l’Introduction au présent document).

Étant donné qu’il y a cinq acides aminés définis de manière spécifique, à savoir, Ala, Tyr, Leu, Gly et Ser, le paragraphe 7.b) de la norme ST.26 prescrit l’intégration de la séquence dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

En vertu du paragraphe 29, tout acide aminé “other” doit être représenté par le symbole “X”. Dans l’exemple, la séquence contient l’acide aminé Hse à la position 2, qui ne figure pas dans l’annexe I, section 3, tableau 3. Il s’ensuit que Hse est un acide aminé “other” et doit être représenté par le symbole “X”.

X1-X4 sont des positions variantes, chacune d’elles pouvant correspondre à A ou à G. Le symbole ambigu le plus restrictif représentant les acides aminés alternatifs A ou G est “X”. La séquence peut donc être représentée sous la forme suivante :

AXXXXXYLGS (SEQ ID NO : 50)

Comme indiqué dans l’introduction au présent document, l’intégration de toutes séquences particulières de première importance pour la divulgation ou les revendications de l’invention est fortement conseillée.

Étant donné que l’acide aminé Hse ne figure pas dans l’annexe I, section 4, tableau 4, une clé de caractérisation “SITE” et un qualificateur “NOTE” doivent être employés et doivent indiquer le nom complet non abrégé de l’~~H~~homosérine, conformément au paragraphe 30 de la norme ST.26.

Selon le paragraphe 27, puisque X1-X4 représentent un choix entre deux acides aminés seulement, une description supplémentaire s’impose. Le paragraphe 94 indique qu’il conviendrait d’utiliser la clé de caractérisation “VARIANT” avec le qualificateur “NOTE”, qui prendrait la valeur “A ou G”. Selon le paragraphe 34 de la norme ST.26, comme elles sont adjacentes et décrites de manière identique, ces positions peuvent faire l’objet d’une description commune à l’aide de la syntaxe “3...6” en tant que descripteur d’emplacement de l’élément INSDFeature\_location.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~3.a), 7.b), 25 à 27, **29**, 30 34, 66, 70, 71 et 94-95

### Paragraphe 30 – Annotation d’un acide aminé modifié

**Exemple 30-1 – Clé de caractérisation “CARBOHYD”**

Une demande de brevet décrit un polypeptide contenant un acide aminé modifié de manière spécifique, dont une chaîne latérale a fait l’objet d’une glycosylation et qui se caractérise par le fait que la Cys correspondant aux positions 4 et 15 du polypeptide forme une liaison disulfure, conformément à la séquence ci‑après :

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Asn(asialyloligosaccharide)-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Le peptide énuméré contient 17 acides aminés définis de manière spécifique. Il y a 16 acides aminés naturels, dont le neuvième (asparagine) a fait l’objet d’une glycosylation. La séquence doit donc être intégrée dans un listage des séquences comme le prescrit le paragraphe 7.b) de la norme ST.26.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Aux termes du paragraphe 29 de la norme ST.26, un acide aminé modifié devrait être représenté dans la séquence comme l’acide aminé non modifié correspondant chaque fois que possible.

La séquence doit donc être intégrée dans un listage des séquences sous la forme suivante :

LEYCLKRWNETISHCAW (SEQ ID NO : 51)

L’acide aminé modifié doit s’accompagner d’une description supplémentaire. La clé de caractérisation “CARBOHYD” et le qualificateur (obligatoire) du type “NOTE” doivent être employés pour indiquer l’occurrence de la fixation d’une chaîne de sucre (asialyloligosaccharide) sur l’asparagine à la position 9. Le qualificateur du type “NOTE” décrit le type de liaison, par exemple une N-liaison. Le descripteur d’emplacement dans l’élément de l’emplacement de la caractéristique est le numéro de position de résidu de l’asparagine modifiée.

En outre, il y a une liaison disulfure entre les deux résidus de Cys. De ce fait, la clé de caractérisation “DISULFID” est donc utilisée pour décrire une liaison intrachaîne. Les descripteurs d’emplacement dans l’élément de l’emplacement de la caractéristique sont les numéros de position des résidus de Cys liés et l’opérateur d’emplacement “assurant un rôle de jonction”, “join(4,15)”. Le qualificateur du type NOTE n’est pas obligatoire.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~3.a), 7.b), 26, 29, **30** et annexe I, section 7, clé de caractérisation 7.4

### Paragraphe 36 – Séquences contenant des régions comportant un nombre exact de résidus contigus “n” ou “X”

**Exemple 36-1 : Séquence dont une région contient un nombre connu de résidus “X” représentée comme une séquence unique**

LL-100-KYMR

où “-100-”entre les acides aminés leucine et lysine correspond à une région de 100 acides aminés dans la séquence.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Le paragraphe 36 de la norme ST.26 prescrit l’intégration d’une séquence qui contient au moins quatre acides aminés définis de manière spécifique et séparés par une ou plusieurs régions contenant un nombre déterminé de résidus “X”.

La séquence divulguée utilise un symbole non conventionnel, à savoir “-100-.” La définition de “-100-” doit être tirée de l’explication de la séquence donnée dans la divulgation, qui définit ce symbole comme 100 acides aminés entre la leucine et la lysine (voir Introduction au présent document). De ce fait, “-100-” est une région déterminée de résidus “X”. Étant donné que six des 106 acides aminés de la séquence sont définis de manière spécifique, le paragraphe 7.b) de la norme ST.26 prescrit l’intégration de la séquence dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Le symbole non conventionnel “-100-” est représenté comme 100 résidus “X” (car tout symbole utilisé pour représenter un acide aminé est l’équivalent d’un seul résidu). Il s’ensuit qu’une séquence unique de 106 acides aminés de longueur, contenant 100 résidus “X” entre LL et KYMR, doit être intégrée dans un listage des séquences (SEQ ID NO : 52).

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7.b), 26, 27, et **36**

**Exemple 36-2 : Séquence dont plusieurs régions contiennent un nombre ou une série connu de résidus “X” représentée comme une séquence unique**

Lys-z2-Lys-zm-Lys-z3-Lys-zn-Lys-z2-Lys

où z représente tout acide aminé, m=20, n=19-20, z2 signifie que les paires de lysine sont séparées par deux acides aminés quels qu’ils soient et z3 signifie que les paires de lysine sont séparées par trois acides aminés quels qu’ils soient.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La séquence divulguée utilise un symbole non conventionnel, à savoir “z.” Il faut donc consulter la divulgation pour déterminer la définition; “z” est défini comme tout acide aminé (voir l’introduction au présent document). Le symbole conventionnel utilisé pour représenter tout acide aminé est “X”. Compte tenu de la présence de variables “X”, le peptide contient six résidus de lysine qui sont énumérés et définis de manière spécifique, à intégrer dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence utilise un symbole non conventionnel “z”, dont la définition doit être tirée de la divulgation. Étant donné que “z” est défini comme tout acide aminé, le symbole conventionnel est “X.”

Il est préférable d’utiliser le moyen de représentation le plus général, qui est le suivant (voir l’introduction au présent document) :

KXXKXXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXXKXXK (SEQ ID NO : 53)

où zn est égal à 20 “X’s”, une description supplémentaire indiquant que la variable “X” correspondant à la position 30 peut être supprimée.

En lieu et place ou en sus de ce qui précède, la séquence peut être représentée sous la forme suivante :

KXXKXXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXKXXK (SEQ ID NO : 54)

où zn est égal à 19 “X’s”, une description supplémentaire indiquant qu’une variable “X” peut être ajoutée entre les numéros de position 29 et 30.

Conformément au paragraphe 27, “X” sera considéré comme l’équivalent de l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire dans le tableau de caractéristiques. Dès lors, si “X” est destiné à représenter “tout acide aminé”, il doit être annoté au moyen de la clé de caractérisation “VARIANT” et d’un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

Chaque fois que possible, chaque “X” devrait être annoté individuellement. Cependant, une région de résidus “X” contigus ou un grand nombre de résidus “X” dispersés dans l’ensemble de la séquence peuvent être décrits conjointement par la clé de caractérisation “VARIANT” en employant la syntaxe “x...y” pour désigner le descripteur d’emplacement, où x et y sont les positions du premier et du dernier résidu “X”, et par un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~Paragraphes~~26, 27, et **36**

**Exemple 36-3 : Séquence dont plusieurs régions contiennent un nombre ou une série connus de résidus “X” et qui est représentée comme une séquence unique**

K-z2-K-zm-K-z3-K-zn-K-z2-K

où z est tout acide aminé, m=15-25, de préférence 20-22, n=15-25, de préférence 19-20, z2 signifie que les paires de lysine sont séparées par deux acides aminés quels qu’ils soient et z3 signifie que les paires de lysine sont séparées par trois acides aminés quels qu’ils soient.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La séquence de l’exemple utilise un symbole non conventionnel, à savoir “z”. De ce fait, on consulte la divulgation pour déterminer la définition de “z” (voir l’Introduction au présent document). La divulgation définit ce symbole comme tout acide aminé. Le symbole conventionnel utilisé pour représenter cet acide aminé est “X.” Compte tenu de la présence de variables “X”, le peptide contient six résidus de lysine qui sont énumérés et définis de manière spécifique, à intégrer dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence utilise un symbole non conventionnel “z”, dont la définition doit être tirée de la divulgation. Étant donné que “z” est défini comme tout acide aminé, le symbole conventionnel est “X”. Le moyen de représentation à préférer et le plus englobant est le suivant :

KXXKXXXXXXXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXXXXXXXKXXK (SEQ ID NO : 55)

(où m=25 et n=25), une description supplémentaire indiquant que l’on peut supprimer jusqu’à 10 résidus “X” dans chacune des régions “zm” ou “zn”.

Comme indiqué dans l’introduction au présent document, l’intégration de toutes séquences particulières de première importance pour la divulgation ou les revendications de l’invention est fortement conseillée.

La séquence pourrait également être représentée sous la forme suivante :

KXXKXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXKXXK (SEQ ID NO : 56)

(où m=15 et n=15), une description supplémentaire indiquant que l’on peut ajouter jusqu’à 10 résidus “X” dans chacune des régions “zm” ou “zn”.

Toutes autres variantes possibles pourraient également être intégrées.

Conformément au paragraphe 27, “X” sera considéré comme l’équivalent de l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire dans le tableau de caractéristiques. Dès lors, si “X” est destiné à représenter “tout acide aminé”, il doit être annoté au moyen de la clé de caractérisation “VARIANT” et d’un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

Chaque fois que possible, chaque “X” devrait être annoté individuellement. Cependant, une région de résidus “X” contigus ou un grand nombre de résidus “X” dispersés dans l’ensemble de la séquence peuvent être décrits conjointement par la clé de caractérisation “VARIANT” en employant la syntaxe “x..y” pour désigner le descripteur d’emplacement, où x et y sont les positions du premier et du dernier résidu “X”, et par un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

**N.B.** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~ 27, et **36**

### Paragraphe 37 – Séquences dont des régions contiennent un nombre inconnu de résidus contigus “n” ou “X”

**Exemple 37-1 : Une séquence dont des régions contiennent un nombre inconnu de résidus “X” ne doit pas être représentée comme une séquence unique**

Gly-Gly----Gly-Gly-Xaa-Xaa

où le symbole ---- est une brèche non définie dans la séquence, Xaa représente tout acide aminé, et les résidus de glycine et de Xaa sont reliés entre eux par des liaisons peptidiques.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**NON**

Le paragraphe 37 de la norme ST.26 interdit l’intégration d’une séquence qui contient une brèche non définie; l’intégration de l’ensemble de la séquence n’est donc pas prescrite.

En revanche, le paragraphe 37 de la norme ST.26 prescrit l’intégration de toute portion de séquence adjacente à une brèche non définie qui contient au moins quatre acides aminés définis de manière spécifique. Dans l’exemple ci‑dessus, l’intégration de toute portion de séquence adjacente à une brèche non définie n’est pas prescrite, car chaque portion ne contient que deux acides aminés définis de manière spécifique.

**Question 2 : La norme ST.26 autorise‑t‑elle l’intégration de la ou des séquences?**

**NON** – pas l’ensemble de la séquence

**NON** – aucune portion de la séquence

Le paragraphe 37 de la norme ST.26 n’autorise pas l’intégration de l’ensemble de la séquence.

Le paragraphe 8 de la norme ST.26 n’autorise pas l’intégration de l’une ou de l’autre des portions adjacentes à la brèche non définie, car chaque portion ne contient que deux acides aminés définis de manière spécifique.

**Paragraphe(s) pertinent(s) de la norme ST.26** :~~paragraphes~~7.b), 8, 26 et **37**

**Exemple 37-2 : Une séquence dont des régions contiennent un nombre inconnu de résidus “X” ne doit pas être représentée comme une séquence unique**

Gly-Gly----Gly-Gly-Ala-Gly-Xaa-Xaa

où le symbole ---- est une brèche non définie dans la séquence, Xaa représente tout acide aminé, et les résidus de glycine et de Xaa sont reliés entre eux par des liaisons peptidiques.

**Question 1 : La norme ST.26 prescrit-elle l’intégration de la ou des séquences?**

**NON** – pas l’ensemble de la séquence

**OUI** – une portion de la séquence

Le paragraphe 37 de la norme ST.26 interdit l’intégration d’une séquence qui contient une brèche non définie, mais prescrit l’intégration de toute portion de séquence adjacente à une brèche non définie qui contient au moins quatre acides aminés définis de manière spécifique.

Dans l’exemple ci‑dessus, la norme ST.26 ne prescrit pas (et interdit) l’intégration de l’ensemble de la séquence, qui contient une brèche non définie, et de la portion Gly-Gly adjacente à ladite brèche, qui ne contient que deux acides aminés définis de manière spécifique. Toutefois, la norme ST.26 prescrit l’intégration de la portion Gly-Gly-Ala-Gly- Xaa-Xaa adjacente à la brèche non définie, car elle contient au moins quatre acides aminés définis de manière spécifique.

**Question 2 : La norme ST.26 autorise‑t‑elle l’intégration de la ou des séquences?**

**NON** – pas de l’ensemble de la séquence ni de la portion Gly-Gly

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La portion de la séquence adjacente à la brèche non définie qui contient quatre acides aminés définis de manière spécifique doit être représentée sous la forme suivante :

GGAGXX (SEQ ID NO : 57)

La séquence devrait de préférence faire l’objet d’une annotation indiquant que la séquence représentée fait partie d’une séquence plus longue qui contient une brèche non définie en utilisant la clé de caractérisation “SITE”, l’emplacement de caractéristique “1” et le qualificateur du type “NOTE” qui prendrait par exemple la valeur “Ce résidu est relié par une liaison N-terminale à un peptide comportant un Gly-Gly N-terminal et une brèche de longueur non définie.”

Conformément au paragraphe 27, “X” sera considéré comme l’équivalent de l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire dans le tableau de caractéristiques. Dès lors, si “X” est destiné à représenter “tout acide aminé”, il doit être annoté au moyen de la clé de caractérisation “VARIANT” et d’un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

Chaque fois que possible, chaque “X” devrait être annoté individuellement. Cependant, une région de résidus “X” contigus ou un grand nombre de résidus “X” dispersés dans l’ensemble de la séquence peuvent être décrits conjointement par la clé de caractérisation “VARIANT” en employant la syntaxe “x..y” pour désigner le descripteur d’emplacement, où x et y sont les positions du premier et du dernier résidu “X”, et par un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7.b),8, 26, 27, et **37**

### Paragraphe 55 – Séquence de nucléotides contenant à la fois des segments d’ADN et d’ARN

**Exemple 55-1 : Molécule combinée d’ADN et d’ARN**

La séquence d’oligonucléotides suivante est divulguée dans une demande de brevet :

AGACCTTcggagucuccuguugaacagauagucaaaguagauC

où les majuscules représentent des résidus d’ADN et les minuscules des résidus d’ARN.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La séquence divulguée contient plus de 10 nucléotides énumérés et définis de manière spécifique; elle doit donc être intégrée dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence nucléotidique doit être intégrée dans un listage des séquences sous la forme suivante :

agaccttcggagtctcctgttgaacagatagtcaaagtagatc (SEQ ID NO : 92)

A noter que les nucléotides d’uracile doivent être représentés par le symbole “t” dans le listage des séquences.

En vertu du paragraphe 55 de la norme ST.26, si une séquence de nucléotides contient à la fois des fragments d’ADN et d’ARN, elle doit apparaître sous forme de molécule de type “DNA”. La molécule combinée d’ADN et d’ARN doit en outre être décrite à l’aide de la clé de caractérisation “source”, du qualificateur obligatoire “organism”, qui prend la valeur “synthetic construct”, et du qualificateur obligatoire “mol\_type”, qui prend la valeur “other DNA”. Chaque fragment de la séquence doit en outre être décrit par la clé de caractérisation “misc\_feature”, qui indique son emplacement, et par le qualificateur “note”, ce dernier précisant s’il s’agit d’un fragment d’ADN ou d’ARN. La séquence divulguée contient deux segments d’ADN (positions 1 à 7 et 43 des nucléotides) et un segment d’ARN (positions 8 à 42 des nucléotides).

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7, 14, **55**, 56 et 83

### Paragraphe 87 – Clé de caractérisation “CDS”

**Exemple 87-1 : Codage de la séquence de nucléotides et séquence d’acides aminés codée**

Une demande de brevet décrit la séquence nucléotidique ci‑après et sa traduction :

atg acc gga aat aaa cct gaa acc gat gtt tac gaa att tta tga

Met Thr Gly Asn Lys Pro Glu Thr Asp Val Tyr Glu Ile Leu STOP

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI** ~~– la séquence nucléotidique.~~

**La séquence nucléotidique énumérée compte plus de 10 nucléotides définis de manière spécifique.**

**~~OUI~~** ~~– la séquence peptidique.~~ La séquence ~~peptidique~~d’acides aminés énumérée contient plus de quatre acides aminés définis de manière spécifique.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence nucléotidique doit être présentée sous la forme suivante :

atgaccggaaataaacctgaaaccgatgtttacgaaattttatga (SEQ ID NO : 58)

La séquence nucléotidique ~~devrait~~doit s’accompagner d’une description supplémentaire utilisant la clé de caractérisation “CDS”, et l’élément INSDFeature\_location devrait désigner l’ensemble de la séquence, y compris le codon d’arrêt (c’est‑à‑dire les positions 1 à 45). De plus, le qualificateur “translation” devrait être intégré en prenant la valeur “MTGNKPETDVYEIL”. La demande ne divulgue pas le tableau du code génétique qui est appliqué à la traduction (voir annexe 1, section 9, tableau 5). Si le tableau de codes normalisés est appliqué, le qualificateur “transl\_table” n’est pas nécessaire; toutefois, si un tableau du code génétique différent est appliqué, il faut indiquer la valeur appropriée figurant dans le tableau 5 pour le qualificateur “transl\_table”. Enfin, le qualificateur “protein\_id” doit être employé pour indiquer le numéro d’identification de la séquence ~~peptidique~~d’acides aminés traduite.

La séquence ~~peptidique~~’acides aminés.doit être présentée de façon distincte au moyen des codes à une lettre ci‑après, et disposer de son propre numéro d’identification de séquence :

MTGNKPETDVYEIL (SEQ ID NO : 59)

Le STOP qui suit la séquence ~~peptidique~~d’acides aminés énumérée ne doit pas être intégré dans la séquence ~~peptidique~~d’acides aminés à l’occasion du listage des séquences.

**N.B.** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7.a), 7.b), 26, 28, **87**, 88 et 90

**Exemple 87-2 : L’emplacement de la caractéristique s’étend au‑delà de la séquence divulguée**

Une demande de brevet contient la figure suivante, qui révèle une séquence de codage partielle et sa séquence d’acides aminés traduite :

cat cac gca gca gaa tgt gga ttt tgt cct caa caa tgg caa gtt cta 48

His His Ala Ala Glu Cys Gly Phe Cys Pro Gln Gln Trp Gln Val Leu

1 5 10 15

cgt ggg agt ctg tgc att tgt gag ggt cca gct gaa gga tgg ttc ata 96

Arg Gly Ser Leu Cys Ile Cys Glu Gly Pro Ala Glu Gly Trp Phe Ile

20 25 30

tca aga tgt tgg tta tgg tgt ggg cct caa gtc caa ggc ttt atc ttt 144

Ser Arg Cys Trp Leu Trp Cys Gly Pro Gln Val Gln Gly Phe Ile Phe

35 40 45

gga gaa ggc aag gaa gga ggc ggt gac aga cgg gct gaa gcg agc cct 192

Gly Glu Gly Lys Glu Gly Gly Gly Asp Arg Arg Ala Glu Ala Ser Pro

50 55 60

cag gag ttt tgg gaa tgc act tgg 216

Gln Glu Phe Trp Glu Cys Thr Trp

65 70

Figure 1 – Séquence de codage partielle du gène *ITCH1* de l’*Homo sapi*ens, qui code les acides aminés 20 à 91 de la protéine ITCH1 comportant 442 acides aminés.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La demande divulgue une séquence nucléotidique et sa séquence d’acides aminés traduite. Contenant plus de 10 nucléotides définis de manière spécifique, la séquence nucléotidique énumérée doit être intégrée dans un listage des séquences.

La séquence d’acides aminés contient plus de 4 acides aminés définis de manière spécifique et doit donc être intégrée dans un listage des séquences en tant que séquence distincte disposant de son propre numéro d’identification.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence nucléotidique doit être intégrée dans un listage des séquences sous la forme suivante :

catcacgcagcagaatgtggattttgtcctcaacaatggcaagttctacgtgggagtctgtgcatttgtgagggtccagctgaaggatggttcatatcaagatgttggttatggtgtgggcctcaagtccaaggctttatctttggagaaggcaaggaaggaggcggtgacagacgggctgaagcgagccctcaggagttttgggaatgcacttgg (SEQ ID NO : 93)

La séquence nucléotidique devrait être décrite plus en détail au moyen d’une clé de caractérisation “CDS”. L’élément INSDFeature\_location doit indiquer l’emplacement de la caractéristique “CDS” dans la séquence, y compris le codon d’arrêt.

La figure décrit une séquence de codage partielle qui ne contient ni le codon de départ ni le codon d’arrêt. Cependant, la description de la séquence indique que le codon de départ se trouve en amont du nucléotide en position 1 et que le codon d’arrêt se trouve en aval de dernier nucléotide en position 216.

La norme ST.26 prescrit que le descripteur d’emplacement ne doit pas comporter de numéros de résidus en dehors de la série indiquée pour la séquence dans l’élément INSDSeq\_sequence. Dès lors, dans l’exemple ci‑dessus, le descripteur d’emplacement de la clé de caractérisation “CDS” ne peut comporter de numéros de position au‑delà de la série 1 à 216. L’emplacement du codon d’arrêt dans l’élément INSDFeature\_location doit être représenté par le symbole ">" pour indiquer que le codon d’arrêt se trouve en aval de la position 216. De même, on peut employer le symbole "<" pour indiquer que le codon de départ se trouve en amont de la position 1. Le descripteur d’emplacement de la clé de caractérisation “CDS” devrait donc apparaître ainsi :

<1..>216

A noter que les symboles "<" et ">" sont des caractères réservés et seront respectivement remplacés par "&lt;" et "&gt;" dans l’instance XML du listage des séquences.

Le qualificateur de “traduction” devrait être intégré en prenant pour valeur la séquence d’acides aminés de la protéine. La figure ne divulgue pas le tableau de codes génétiques s’appliquant à la traduction (voir l’annexe 1, section 9, tableau 5). Si le tableau de codes normalisés est appliqué, le qualificateur “transl\_table” n’est pas nécessaire; toutefois, si un tableau du code génétique différent est appliqué, il faut indiquer la valeur appropriée figurant dans le tableau 5 de l’Annexe I pour le qualificateur “transl\_table”. Enfin, le qualificateur “protein \_id” doit être intégré dans la caractéristique “CDS”, dont le qualificateur doit indiquer le numéro d’identification de la séquence d’acides aminés traduite.

La séquence d’acides aminés traduite doit apparaître en tant que séquence distincte et disposer de son propre numéro d’identification :

HHAAECGFCPQQWQVLRGSLCICEGPAEGWFISRCWLWCGPQVQGFIFGEGKEGGGDRRAEASPQEFWECTW (SEQ ID NO : 94)

**N.B.** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7, 41, 65, 66, 70, 71, **87** et 90

### Paragraphe 90 – Séquence d’acides aminés codée selon une séquence de codage

**Exemple 90-1 : Séquence d’acides aminés codée selon une séquence de codage avec introns**

Une demande de brevet contient la figure ci‑après divulguant une séquence de codage et sa traduction :



Figure 1 – Les nucléotides apparaissant en gras sont des régions appelées introns.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La demande divulgue une séquence nucléotidique et son passage à une séquence d’acides aminés (traduction). Contenant plus de 10 nucléotides définis de manière spécifique, la séquence nucléotidique énumérée doit être intégrée dans un listage des séquences comme une séquence unique.

La séquence nucléotidique contient des séquences codantes (exons) séparées par des séquences non codantes (introns). La figure décrit la traduction de la séquence nucléotidique sous la forme de trois séquences d’acides aminés non contiguës. Selon la légende de la figure, les régions de nucléotides apparaissant en gras sont des séquences appelées introns qui, après épissage d’un ARN transcrit, seront traduits en protéines. Il s’ensuit que les trois séquences d’acides aminés sont en fait une séquence énumérée contiguë unique qui contient plus de quatre acides aminés définis de manière spécifique et doit être intégrée dans un listage des séquences comme séquence unique.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence nucléotidique doit être intégrée dans un listage des séquences sous la forme suivante :

atgaagactttcgcagccttgctttccgctgtcactctcgcgctctcggtgcgcgcccaggcggctgtctggagtcaatgtaagtgccgctgcttttcattgatacgagactctacgccgagctgacgtgctaccgtataggtggcggtacaccgggttggacgggcgagaccacttgcgttgctggttcggtttgtacctccttgagctcagtgagcgactttcaatccgtcgtcattgctcctcatgtattgacgattggccttcatagtcatactctcaatgcgttccgggctccgcaacgtccagcgctccggcggccccctcagcgacaacttcaggccccgcacctacggacggaacgtgctcggccagcggggcatggccgccattgacctga (SEQ ID NO : 74)

La séquence nucléotidique ~~devrait~~doit s’accompagner d’une description supplémentaire mettant en jeu une clé de caractérisation “CDS”, et l’élément INSDFeature\_location devrait désigner l’emplacement de la séquence codante, y compris le codon d’arrêt indiqué par “Ter”. De plus, le qualificateur de “traduction” devrait être intégré en prenant pour valeur la séquence d’acides aminés de la protéine. (On notera que le symbole de fin “Ter” à la dernière position de la séquence ne doit pas être intégré dans la séquence d’acides aminés.) La demande ne divulgue pas le tableau du code génétique qui est appliqué à la traduction (voir annexe 1, section 9, tableau 5). Si le tableau de codes normalisés est appliqué, le qualificateur “transl\_table” n’est pas nécessaire; toutefois, si un tableau du code génétique différent est appliqué, il faut indiquer la valeur appropriée figurant dans le tableau 5 pour le qualificateur “transl\_table”. Enfin, il faut intégrer le qualificateur “protein \_id” dont la valeur indiquera le numéro d’identification de séquence de la séquence d’acides aminés traduite.

La séquence d’acides aminés doit être intégrée comme une séquence unique :

MKTFAALLSAVTLALSVRAQAAVWSQCGGTPGWTGETTCVAGSVCTSLSSSYSQCVPGSATSSAPAAPSATTSGPAPTDGTCSASGAWPPLT (SEQ ID NO : 75)

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7, 26, 28, 57, et 87 à **90**

### Paragraphe 91 – Paragraphe 91 – Séquence primaire et une variante, chacune énumérée par son résidu

**Exemple 91-1 : Représentation des variantes énumérées**

La description comprend l’alignement de séquence ci‑après.

*D. melanogaster ACATTGAATCTCATACCACTTT*

*D. virilis ...-..G...C..--.G.....*

*D. simulans GT..G.CG..GT..SGT.G...*

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Il est courant dans la branche d’inclure des “points” dans un alignement de séquence pour indiquer que “cette position est la même que la position précédente”. On considère que les points dans les séquences de *~~species 2~~ D~~.~~virilis* et de *~~3~~ D~~.~~simulans* correspondent à des nucléotides énumérés et définis de manière spécifique; c’est un moyen simple d’indiquer qu’une position donnée est occupée par le même nucléotide que dans la séquence de *~~species 1~~D. melanog~~f~~aster.* De plus, les alignements de séquence présentent souvent le symbole “-” pour indiquer l’absence d’un résidu afin de maximiser l’alignement.

Les séquences de nucléotides de *D*. *melanog~~f~~aster* ~~species 1~~ et de *D*. *simulans*~~3~~ contiennent donc vingt‑deux nucléotides énumérés et définis de manière spécifique, tandis que la séquence nucléotidique de ~~species 2~~*D. virilis* en contient dix‑neuf. De ce fait, chaque séquence doit, en vertu du paragraphe 7.a) de la norme ST.26, être intégrée dans un listage des séquences et disposer d’un numéro d’identification de séquence distinct.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence Drosophila melanogaster doit être intégrée dans un listage des séquences sous la forme suivante :

acattgaatctcataccacttt (SEQ ID NO : 60)

La séquence Drosophila virilis doit être intégrée dans un listage des séquences sous la forme suivante :

acatggatcccacgacttt (SEQ ID NO : 61)

La séquence Drosophila simulans doit être intégrée dans un listage des séquences sous la forme suivante :

gtatggcgtcgtatsgtagttt (SEQ ID NO : 62)

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7.a), 13 et **91**

**Exemple 91-2 : Représentation des variantes énumérées**

La description comprend le tableau ci‑après d’un peptide et de ses variantes fonctionnelles. Dans ce tableau, un espace blanc indique qu’un acide aminé de la variante est le même que l’acide aminé correspondant de la “Séquence” et un “-” indique une suppression de l’acide aminé correspondant de la “Séquence”.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Position | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Séquence | A | V | L | t | y | L | r | g | E |
| Variante 1 |  |  |  |  |  |  |  |  | A |
| Variante 2 |  |  | P |  |  | P |  |  |  |
| Variante 3 |  |  | A | I | g | y |  |  |  |
| Variante 4 |  |  |  |  |  |  | - |  |  |

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Comme indiqué, un espace blanc dans le tableau indique qu’un acide aminé de la variante est le même que l’acide aminé correspondant de la “Séquence”. De ce fait, les acides aminés des séquences variantes sont énumérés et définis de manière spécifique.

Étant donné que les quatre séquences variantes contiennent chacune plus de quatre acides aminés énumérés et définis de manière spécifique, chaque séquence doit, en vertu du paragraphe 7.~~a~~b) de la norme ST.26, être intégrée dans un listage des séquences et disposer d’un numéro d’identification de séquence distinct.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

AVLTYLRGE (SEQ ID NO : 76)

AVLTYLRGA (SEQ ID NO : 77)

AVPTYPRGE (SEQ ID NO : 78)

AVAIGYRGE (SEQ ID NO : 79)

AVLTYLGE (SEQ ID NO : 80)

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7.b), 26, et **91**

**Exemple 91-3 : Représentation d’une séquence consensus**

Une demande de brevet comprend la figure 1 présentant l’alignement séquentiel multiple ci‑après.

*Consensus* LEG**n**EQFINA**ak**IIRHP**k**Y**nrk**T**ln**NDI**m**LIK

*Homo sapiens* LEG**N**EQFINA**AK**IIRHP**Q**Y**DRK**T**LN**NDI**M**LIK

*Pongo abelii* LEG**N**EQFINA**AK**IIRHP**Q**Y**DRK**T**VN**NDI**M**LIK

*Papio Anubis* LEG**T**EQFINA**AK**IIRHP**D**Y**DRK**T**LN**NDI**L**LIK

*Rhinopithecus roxellana* LEG**T**EQFINA**AK**IIRHP**N**Y**NRI**T**LD**NDI**L**LIK

*Pan paniscus* LEG**N**EQFINA**AK**IIRHP**K**Y**NRI**T**LN**NDI**M**LIK

*Rhinopithecus bieti* LEG**N**EQFINA**TK**IIRHP**K**Y**NGN**T**LN**NDI**M**LIK

*Rhinopithecus roxellana* LEG**N**EQFINA**TQ**IIRHP**K**Y**NGN**T**LN**NDI**M**LIK

La séquence consensus comprend des lettres majuscules qui représentent les résidus d’acides aminés conservés, tandis que les lettres minuscules “n”, “a”, “k”, “r”, “l” et “m” représentent les résidus d’acides aminés prédominants parmi les séquences alignées.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

**Les lettres minuscules de la séquence consensus représentent chacune un résidu d’acides aminés unique. En conséquence, la séquence consensus contient, comme chacune des sept autres séquences de la figure 1, au moins quatre acides aminés définis de manière spécifique. Le paragraphe 7.b) de la norme ST.26 prescrit l’intégration des huit séquences dans le listage des séquences.**

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Dans la séquence consensus, les lettres minuscules sont utilisées comme symboles ambigus pour représenter l’acide aminé prédominant parmi les variantes possibles d’une position particulière. De ce fait, les lettres minuscules “n”, “a”, “k”, “r”, “l” et “m” sont des symboles conventionnels utilisés d’une manière non conventionnelle, et la séquence consensus doit être représentée au moyen d’un symbole ambigu à la place de chacune des lettres minuscules.

Il faudrait utiliser le symbole ambigu le plus restrictif. Pour la plupart des positions de la séquence consensus, “X” est le symbole ambigu le plus restrictif; toutefois, le symbole ambigu le plus restrictif pour “D” ou “N” aux positions 20 et 25 est “B”. La séquence consensus doit être intégrée dans le listage des séquences sous la forme suivante :

LEGXEQFINAXXIIRHPXYBXXTXBNDIXLIK (SEQ ID NO : 81)

Conformément au paragraphe 27, le symbole “X” sera considéré comme l’équivalent de l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire dans le tableau de caractéristiques. De ce fait, chaque “X” de la séquence consensus doit s’accompagner d’une description supplémentaire dans un tableau de caractéristiques au moyen de la clé de caractérisation “VARIANT” et du qualificateur du type “NOTE” pour indiquer les variantes possibles de chaque position.

Les sept autres séquences doivent être intégrées dans le listage des séquences sous la forme suivante :

LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK (SEQ ID NO : 82)

LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK (SEQ ID NO : 83)

LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK (SEQ ID NO : 84)

LEGTEQFINAAKIIRHPNYNRITLDNDILLIK (SEQ ID NO : 85)

LEGNEQFINAAKIIRHPKYNRITLNNDIMLIK (SEQ ID NO : 86)

LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK (SEQ ID NO : 87)

LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK (SEQ ID NO : 88)

**N.B.** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7.b), 26, 27, **91** et 95

### Paragraphe 92 – Séquence variante divulguée comme une séquence unique avec des résidus alternatifs énumérés

**Exemple 92-1 : Représentation d’une séquence unique avec des acides aminés alternatifs énumérés**

Une demande de brevet est déposée au titre d’un peptide dont la séquence est la suivante :

(i) Gly-Gly-Gly-[Leu ou Ile]-Ala-Thr-[Ser ou Thr]

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

La séquence contenant quatre acides aminés définis de manière spécifique, le paragraphe 7.b) de la norme ST.26 prescrit son intégration dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Le tableau 3 de l’annexe I, section 3 définit le symbole ambigu “J” comme désignant l’isoleucine ou la leucine. De ce fait, la représentation à préférer de la séquence est la suivante :

GGGJATX (SEQ ID NO : 63)

qui doit s’accompagner d’une description supplémentaire dans un tableau de caractéristiques réalisée au moyen de la clé de caractérisation “VARIANT” et du qualificateur du type “NOTE” pour indiquer que “X” est la ~~S~~sérine ou la ~~T~~thréonine.

Une autre solution consiste à représenter la séquence, par exemple, de la façon suivante :

GGGLATS (SEQ ID NO : 64)

qui doit s’accompagner d’une description supplémentaire dans un tableau de caractéristiques réalisée au moyen de la clé de caractérisation “VARIANT” et du qualificateur du type “NOTE” pour indiquer que L peut être remplacé par I, et S par T.

**N.B.  :** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7.b), 26, 27, **92** et 95

### Paragraphe 93.a) – Toute séquence variante divulguée uniquement par référence à une séquence primaire comportant plusieurs variations indépendantes

**Exemple 93.a)-1 : Représentation d’une séquence variante par annotation de la séquence primaire**

Une demande contient la divulgation ci‑après :

“Le fragment peptidique 1 est Gly-Leu-Pro-Xaa-Arg-Ile-Cys, où Xaa peut représenter tout acide aminé…

Dans un autre mode de réalisation, le fragment peptidique 1 est Gly-Leu-Pro-Xaa-Arg-Ile-Cys, où Xaa peut représenter Val, Thr ou Asp….

Dans un autre mode de réalisation, le fragment peptidique 1 est Gly-Leu-Pro-Xaa-Arg-Ile-Cys, où Xaa peut représenter Val.”

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Dans chacun des trois modes de réalisation divulgués, le “fragment peptidique 1” contient au moins six acides aminés définis de manière spécifique; de ce fait, la séquence doit être intégrée dans un listage des séquences, comme le prescrit le paragraphe 7.b) de la norme ST.26.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

Dans cet exemple, la séquence énumérée du “fragment peptidique 1” est divulguée trois fois, en tant que trois modes de réalisation différents, chacun donnant lieu à une description différente de Xaa. Dans cet exemple, “X” est le symbole ambigu le plus restrictif pour la position Xaa.

La norme ST.26 prescrit d’intégrer une seule fois la séquence énumérée divulguée. Dans le plus englobant des trois modes de réalisation, Xaa représente tout acide aminé (voir l’Introduction au présent document). De ce fait, la séquence à intégrer dans le listage des séquences est la suivante :

GLPXRIC (SEQ ID NO : 65)

Conformément au paragraphe 27, “X” sera considéré comme l’équivalent de l’un des symboles “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” ou “V”, sauf s’il est accompagné d’une description supplémentaire dans le tableau de caractéristiques. Dès lors, si “X” est destiné à représenter “tout acide aminé”, il doit être annoté au moyen de la clé de caractérisation “VARIANT” et d’un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

Chaque fois que possible, chaque “X” devrait être annoté individuellement. Cependant, une région de résidus “X” contigus ou un grand nombre de résidus “X” dispersés dans l’ensemble de la séquence peuvent être décrits conjointement par la clé de caractérisation “VARIANT” en employant la syntaxe “x..y” pour désigner le descripteur d’emplacement, où x et y sont les positions du premier et du dernier résidu “X”, et par un qualificateur “NOTE” prenant la valeur “X peut être tout acide aminé”.

Comme indiqué dans l’introduction au présent document, il est fortement conseillé d’intégrer toutes séquences supplémentaires de première importance pour la divulgation ou les revendications de l’invention.

Pour l’exemple ci‑dessus, il est vivement conseillé d’intégrer dans le listage des séquences les trois séquences supplémentaires ci‑après et d’attribuer à chacune son propre numéro d’identification de séquence :

GLPVRIC (SEQ ID NO : 66)

GLPTRIC (SEQ ID NO : 67)

GLPDRIC (SEQ ID NO : 68)

**N.B.** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7.b), 26, 27, et **93a)**

### Paragraphe 93.b) – Toute séquence variante divulguée uniquement par référence à une séquence primaire comportant plusieurs variations interdépendantes

**Exemple 93.b)‑1 : Représentation des séquences variantes individuelles comportant plusieurs variations interdépendantes**

Une demande de brevet décrit la séquence consensus ci‑après :

cgaatgn1cccactacgaatgn2cacgaatgn3cccaca

où n1, n2, et n3 peuvent être a, t, g ou c.

Plusieurs séquences variantes sont divulguées comme suit :

si n1 représente a, n2 et n3 représentent t, g ou c;

si n1 représente t, n2 et n3 représentent a, g ou c;

si n1 représente g, n2 et n3 représentent t, a ou c;

si n1 représente c, n2 et n3 représentent t, g ou a.

**Question 1 : La norme ST.26 prescrit‑elle l’intégration de la ou des séquences?**

**OUI**

Contenant plus de 10 nucléotides énumérés et “définis de manière spécifique”, la séquence doit, en vertu du paragraphe 7.a) de la norme ST.26, être intégrée dans un listage des séquences.

**Question 3 : Comment la ou les séquences devraient‑elles être représentées dans le listage des séquences?**

La séquence énumérée contient plus de 10 nucléotides définis de manière spécifique et trois résidus “n”. La norme ST.26 prescrit l’intégration de la séquence énumérée divulguée et, lorsqu’il convient d’employer un symbole ambigu, il faut choisir le plus restrictif. Dans cet exemple, n1, n2, et n3 peuvent représenter a, t, g ou c, si bien que “n” est le symbole ambigu le plus restrictif. La séquence à intégrer dans le listage des séquences est donc la suivante :

cgaatgncccactacgaatgncacgaatgncccaca (SEQ ID NO : 69)

La séquence énumérée comporte des variations à trois emplacements distincts et les occurrences de ces variations sont interdépendantes. Comme indiqué dans l’introduction au présent document, il est fortement conseillé d’intégrer les séquences supplémentaires qui représentent des modes de réalisation supplémentaires constituant une partie essentielle de l’invention. De ce fait, et conformément au paragraphe 93.b) de la norme ST.26, les modes de réalisation supplémentaires devraient être intégrés dans un listage des séquences en tant que quatre séquences distinctes, chacune disposant de son propre numéro d’identification de séquence :

cgaatgacccactacgaatgbcacgaatgbcccaca (SEQ ID NO : 70)

cgaatgtcccactacgaatgvcacgaatgvcccaca (SEQ ID NO : 71)

cgaatggcccactacgaatghcacgaatghcccaca (SEQ ID NO : 72)

cgaatgccccactacgaatgdcacgaatgdcccaca (SEQ ID NO : 73)

(On notera que b = t, g ou c; v = a, g ou c; h = t, a ou c; et d = t, g ou a; voir annexe I, section 1, tableau 1)

En vertu du paragraphe 15 de la norme ST.26, il faut utiliser le symbole le plus restrictif pour représenter les positions variables. Il s’ensuit que n2 et n3 ne doivent pas être représentés par “n” dans la séquence.

**N.B.** La représentation préférée de la séquence indiquée ci‑dessus sert à fournir un listage des séquences à la date du dépôt d’une demande de brevet. La même représentation pourra ne pas être applicable à un listage des séquences fourni après cette date, car il faut tenir compte de la question de savoir si l’information fournie pourrait être prise en considération par un office de la propriété intellectuelle pour ajouter des éléments à la divulgation originale.

**Paragraphe(s) pertinent(s) de la norme ST.26 :** ~~paragraphes~~7.a), 15, et **93b)**

[L’appendice de l’annexe VI de la norme ST.26 suit]

## **APPENDIX**

GUIDANCE DOCUMENT SEQUENCES IN XML

<?xml version="1.0" encoding="UTF-8"?>

<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.~~0~~2//EN" "~~resources/~~ST26SequenceListing\_V1\_~~0~~2.dtd">

<ST26SequenceListing dtdVersion="1.2" fileName="Guidance\_Document\_Sequences\_XML" softwareName="prototype" softwareVersion="1.0" productionDate="~~2017-01-02~~2018-07-09">

<ApplicantFileReference>ABCD#1234</ApplicantFileReference>

<ApplicantName languageCode="de">Johannes ~~Jäger~~JÃ¤ger</ApplicantName>

<ApplicantNameLatin>Johannes Jaeger</ApplicantNameLatin>

<InventionTitle languageCode="de">Pharmakologische Wirkstoffe ~~für~~fÃ¼r das Nervensystem</InventionTitle>

<SequenceTotalQuantity>~~88~~94</SequenceTotalQuantity>

<SequenceData sequenceIDNumber="1">

<INSDSeq>

<INSDSeq\_length>7</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>REGION</INSDFeature\_key>

<INSDFeature\_location>1..7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>circular peptide</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>1</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>D-Alanine</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>2</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>D-Glutamic acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>Nle</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>D-Methionine</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>D-Norleucine</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>AEKXGMX</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="2">

<INSDSeq>

<INSDSeq\_length>6</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>1</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>the N-terminus is acetylated</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>6-amino-7-(1H-indol-3-yl)-5-oxoheptanoic acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>the C-terminus is methylated</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>VAFXGK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="3">

<INSDSeq>

<INSDSeq\_length>4</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>1</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>the N-terminus is acetylated</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>C-terminus linked via a glutaraldehyde bridge to dipeptide GK</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>VAFW</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="4">

<INSDSeq>

<INSDSeq\_length>5</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>REGION</INSDFeature\_key>

<INSDFeature\_location>&gt;5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>The entire sequence of amino acids 1-5 can be repeated one or more times</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GGGGX</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="5">

<INSDSeq>

<INSDSeq\_length>12</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..12</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>misc\_feature</INSDFeature\_key>

<INSDFeature\_location>12</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>cytosine attached to a C3 spacer, which is joined to another nucleic acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>atgcatgcatgc</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="6">

<INSDSeq>

<INSDSeq\_length>12</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..12</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>misc\_feature</INSDFeature\_key>

<INSDFeature\_location>1</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>cytosine attached to a C3 spacer, which is joined to another nucleic acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>cggcatgcatgc</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="7">

<INSDSeq>

<INSDSeq\_length>25</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..25</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Mus musculus</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>atgcatgcatgcvcggcatgcatgc</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="8">

<INSDSeq>

<INSDSeq\_length>12</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..12</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>misc\_feature</INSDFeature\_key>

<INSDFeature\_location>12</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>cytosine is linked to a C3 spacer, which is linked to 5&amp;apos;-end of another nucleic acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>atgcatgcatgc</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="9">

<INSDSeq>

<INSDSeq\_length>12</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..12</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>misc\_feature</INSDFeature\_key>

<INSDFeature\_location>1</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>cytosine is linked to a C3 spacer, which is linked to 3&amp;apos;-end of another nucleic acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>cggcatgcatgc</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="10">

<INSDSeq>

<INSDSeq\_length>18</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..18</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>modified\_base</INSDFeature\_key>

<INSDFeature\_location>11</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>mod\_base</INSDQualifier\_name>

<INSDQualifier\_value>OTHER</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>abasic site</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>gagcattgacntaaggct</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="11">

<INSDSeq>

<INSDSeq\_length>30</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..30</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>GNA sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>modified\_base</INSDFeature\_key>

<INSDFeature\_location>1..30</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>mod\_base</INSDQualifier\_name>

<INSDQualifier\_value>OTHER</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value> 2,3-dihydroxypropyl nucleosides (glycol nucleic acids)</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>tagttcattgactaaggctccccattgact</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="12">

<INSDSeq>

<INSDSeq\_length>10</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..10</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>aatgccggag</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="13">

<INSDSeq>

<INSDSeq\_length>10</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..10</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>aatgccggak</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="14">

<INSDSeq>

<INSDSeq\_length>10</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..10</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>aatgccggac</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="15">

<INSDSeq>

<INSDSeq\_length>10</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..10</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>aatgttggac</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="16">

<INSDSeq>

<INSDSeq\_length>15</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..15</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>nngkngkngkagvcr</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="17">

<INSDSeq>

<INSDSeq\_length>7</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>1</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>XYEKGJL</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="18">

<INSDSeq>

<INSDSeq\_length>30</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..30</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>misc\_feature</INSDFeature\_key>

<INSDFeature\_location>1</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>This sequence is one of four branches of a branched polynucleotide</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>cacacaaaaaaaaaaaaaaaaaaaaaaaaa</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="19">

<INSDSeq>

<INSDSeq\_length>28</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..28</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>misc\_feature</INSDFeature\_key>

<INSDFeature\_location>1</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>This sequence is one of four branches of a branched polynucleotide.</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>cacataggcatctcctagtgcaggaaga</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="20">

<INSDSeq>

<INSDSeq\_length>75</INSDSeq\_length>

<INSDSeq\_moltype>RNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..75</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>tRNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>modified\_base</INSDFeature\_key>

<INSDFeature\_location>39</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>mod\_base</INSDQualifier\_name>

<INSDQualifier\_value>p</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>modified\_base</INSDFeature\_key>

<INSDFeature\_location>54</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>mod\_base</INSDQualifier\_name>

<INSDQualifier\_value>p</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>gcggatttagctcagctgggagagcgccagactgaatanctggagtcctgtgtncgatccacagaattcgcacca</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="21">

<INSDSeq>

<INSDSeq\_length>44</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..44</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>gatcattttttatattttttatattttttatattttttatgtac</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="22">

<INSDSeq>

<INSDSeq\_length>44</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..44</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>repeat\_region</INSDFeature\_key>

<INSDFeature\_location>5..40</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>rpt\_type</INSDQualifier\_name>

<INSDQualifier\_value>tandem</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>rpt\_unit\_seq</INSDQualifier\_name>

<INSDQualifier\_value>atacgcact</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>attcatacgcactatacgcactatacgcactatacgcactgtac</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="23">

<INSDSeq>

<INSDSeq\_length>44</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..44</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>repeat\_region</INSDFeature\_key>

<INSDFeature\_location>5..40</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>rpt\_type</INSDQualifier\_name>

<INSDQualifier\_value>tandem</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>rpt\_unit\_seq</INSDQualifier\_name>

<INSDQualifier\_value>atacgcact</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>gatcatacgcactatacgcactatacgcactatacgcactgtac</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="24">

<INSDSeq>

<INSDSeq\_length>12</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..12</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>misc\_difference</INSDFeature\_key>

<INSDFeature\_location>5..8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>A nucleotide in any of positions 5-8 may be replaced with inosine or pseudouridine.</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>modified\_base</INSDFeature\_key>

<INSDFeature\_location>5..8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>mod\_base</INSDQualifier\_name>

<INSDQualifier\_value>OTHER</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>i or p</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>gatcaaaagtac</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="25">

<INSDSeq>

<INSDSeq\_length>74</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..74</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>1..8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>10..19</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>21..48</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>50..68</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>70..74</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>XXXXXXXXDXXXXXXXXXXFXXXXXXXXXXXXXXXXXXXXXXXXXXXXAXXXXXXXXXXXXXXXXXXXGXXXXX</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="26">

<INSDSeq>

<INSDSeq\_length>4</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>1..4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>This sequence is one part of a branched amino acid sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>Leu is linked to another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>RISL</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="27">

<INSDSeq>

<INSDSeq\_length>4</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>1..4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>This sequence is one part of a branched ~~peptide~~amino acid sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>3</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>The side chain of Lysine is linked via an amide linkage to another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>The side chain of Lysine is linked via an amide linkage to another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LLKK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="28">

<INSDSeq>

<INSDSeq\_length>6</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>1..6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>This sequence is one part of a branched ~~peptide~~amino acid sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>Ala is linked to another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>IPACTA</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="29">

<INSDSeq>

<INSDSeq\_length>6</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>1..6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>This sequence is one part of a branched ~~peptide~~amino acid sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>The side chain of Lysine is linked via an amide linkage to another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>FRAGGK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="30">

<INSDSeq>

<INSDSeq\_length>5</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>1..5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>This sequence is one part of a branched ~~peptide~~amino acid sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>Ala is linked to another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>HQYFA</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="31">

<INSDSeq>

<INSDSeq\_length>8</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>1..8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>This sequence is one part of a branched ~~peptide~~amino acid sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>The side chain of Lysine is linked via an amide linkage to another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>The side chain of Lysine is linked via an amide linkage to another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>The side chain of Lysine is linked via an amide linkage to another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>ATFGKKKA</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="32">

<INSDSeq>

<INSDSeq\_length>~~7~~8</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..~~7~~8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>~~Bound~~The side chain of Lysine is linked via an amide linkage ~~to Cly at position 5 of AASHG~~another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>~~DCSAKKK~~DGSAKKKK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="33">

<INSDSeq>

<INSDSeq\_length>5</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>~~Bound~~Gly is linked to ~~Lys at position 5 of DCSAKKK~~another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>AASHG</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="34">

<INSDSeq>

<INSDSeq\_length>13</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..13</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>atcgcaattggcc</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="35">

<INSDSeq>

<INSDSeq\_length>13</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..13</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>ggccaattgcgat</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="36">

<INSDSeq>

<INSDSeq\_length>58</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..58</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="37">

<INSDSeq>

<INSDSeq\_length>16</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..16</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>modified\_base</INSDFeature\_key>

<INSDFeature\_location>~~6~~1..16</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>mod\_base</INSDQualifier\_name>

<INSDQualifier\_value>OTHER</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>N-(2-aminoethyl) glycine nucleosides (PNA)</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>modified\_base</INSDFeature\_key>

<INSDFeature\_location>6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>mod\_base</INSDQualifier\_name>

<INSDQualifier\_value>OTHER</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>N-(2-aminoethyl) glycine 5-nitroindole or N-(2-aminoethyl) glycine 3-nitroindole</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

~~<INSDFeature>~~

~~<INSDFeature\_key>modified\_base</INSDFeature\_key>~~

~~<INSDFeature\_location>1..16</INSDFeature\_location>~~

~~<INSDFeature\_quals>~~

~~<INSDQualifier> <INSDQualifier\_name>mod\_base</INSDQualifier\_name> <INSDQualifier\_value>OTHER</INSDQualifier\_value>~~

~~</INSDQualifier>~~

~~<INSDQualifier> <INSDQualifier\_name>note</INSDQualifier\_name>~~

~~<INSDQualifier\_value>N-(2-aminoethyl) glycine nucleosides (PNA)</INSDQualifier\_value>~~

~~</INSDQualifier>~~

~~</INSDFeature\_quals>~~

~~</INSDFeature>~~

</INSDSeq\_feature-table>

<INSDSeq\_sequence>cgcctnagtcaatggg</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="38">

<INSDSeq>

<INSDSeq\_length>25</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..25</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>atcgggatcgcatattcgattggcc</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="39">

<INSDSeq>

<INSDSeq\_length>25</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..25</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>ggccaatatggcttgcgatcccgat</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="40">

<INSDSeq>

<INSDSeq\_length>20</INSDSeq\_length>

<INSDSeq\_moltype>RNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..20</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>misc\_feature</INSDFeature\_key>

<INSDFeature\_location>1</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>~~ccugucgt is attached at its 3&amp;apos;-end to a linker which is attached to the~~ The 5~~&amp;~~apos; oxygen of the thymidine~~. The~~ is attached through the linker ~~is~~(4-(3-hydroxybenzamido)butyl) phosphinic acid to ccugucgt</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>modified\_base</INSDFeature\_key>

<INSDFeature\_location>1</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>mod\_base</INSDQualifier\_name>

<INSDQualifier\_value>OTHER</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>thymine</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>tcctgtccggagatgttgat</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="41">

<INSDSeq>

<INSDSeq\_length>8</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GGGXGGGX</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="42">

<INSDSeq>

<INSDSeq\_length>400</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..400</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>~~9~~4..400</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>9..400</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>Up to 98 copies of GGGX can be deleted</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGXGGGX</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="43">

<INSDSeq>

<INSDSeq\_length/>

<INSDSeq\_moltype/>

<INSDSeq\_division/>

<INSDSeq\_sequence>000</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="44">

<INSDSeq>

<INSDSeq\_length>8</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GGGXGGGX</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="45">

<INSDSeq>

<INSDSeq\_length>12</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..12</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>12</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GGGXGGGXGGGX</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="46">

<INSDSeq>

<INSDSeq\_length>306</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..306</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>CDS</INSDFeature\_key>

<INSDFeature\_location>19..114</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>translation</INSDQualifier\_name>

<INSDQualifier\_value>MAPNTQTASPRALADSLMQLARQVSRLESGQ</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>protein\_id</INSDQualifier\_name>

<INSDQualifier\_value>47</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>CDS</INSDFeature\_key>

<INSDFeature\_location>115..222</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>translation</INSDQualifier\_name>

<INSDQualifier\_value>MTMITDSLAVVLQRRDWENPGVTQLNRLAAHWCQK</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>protein\_id</INSDQualifier\_name>

<INSDQualifier\_value>48</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>CDS</INSDFeature\_key>

<INSDFeature\_location>251..283</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>translation</INSDQualifier\_name>

<INSDQualifier\_value>MLRRQVNEVA</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>protein\_id</INSDQualifier\_name>

<INSDQualifier\_value>49</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>caattcagggtggtgaatatggcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatggaaagcgggcagtgaatgaccatgattacggattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacattggtgtcaaaaataataataaccggatgtactatttatccctgatgctgcgtcgtcaggtgaatgaagtcgcttaagcaatcaatgtcggatgcggcgcgacgcttatccgaccaacatatcataa</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="47">

<INSDSeq>

<INSDSeq\_length>31</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..31</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>MAPNTQTASPRALADSLMQLARQVSRLESGQ</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="48">

<INSDSeq>

<INSDSeq\_length>35</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..35</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>MTMITDSLAVVLQRRDWENPGVTQLNRLAAHWCQK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="49">

<INSDSeq>

<INSDSeq\_length>10</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..10</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>MLRRQVNEVA</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="50">

<INSDSeq>

<INSDSeq\_length>10</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..10</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>2</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>Homoserine</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>3..6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>A or G</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>AXXXXXYLGS</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="51">

<INSDSeq>

<INSDSeq\_length>17</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..17</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>DISULFID</INSDFeature\_key>

<INSDFeature\_location>join(4,15)</INSDFeature\_location>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>CARBOHYD</INSDFeature\_key>

<INSDFeature\_location>9</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>Asn side-chain N-linked to asialyloligosaccharide</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LEYCLKRWNETISHCAW</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="52">

<INSDSeq>

<INSDSeq\_length>106</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..106</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXKYMR</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="53">

<INSDSeq>

<INSDSeq\_length>53</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..53</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>~~30~~2..3</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>5..24</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>26..28</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>30</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>deleted</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>30..49</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>51..52</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>KXXKXXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXXKXXK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="54">

<INSDSeq>

<INSDSeq\_length>52</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..52</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>~~29^30~~2..3</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>5..24</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>26..28</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>29^30</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X is inserted</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>30..48</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>50..51</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>KXXKXXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXKXXK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="55">

<INSDSeq>

<INSDSeq\_length>63</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..63</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

~~<INSDFeature\_location>5..29</INSDFeature\_location>~~

<INSDFeature\_location>2..3</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>~~Up to 10 residues may~~ X can be ~~deleted~~ any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>~~31..59~~5..29</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>~~Up~~X can be any amino acid; up to 10 residues may be deleted</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>31..33</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>35..59</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid; up to 10 residues may be deleted</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>61..62</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>KXXKXXXXXXXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXXXXXXXKXXK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="56">

<INSDSeq>

<INSDSeq\_length>43</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..43</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

~~<INSDFeature\_location>5..~~ <INSDFeature\_location>2..3</INSDFeature\_location>

~~19</INSDFeature\_location>~~

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>~~Up to 10~~X ~~residues may~~ can be ~~inserted~~any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>~~21..39~~5..19</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>~~Up~~X can be any amino acid; up to 10 X residues may be inserted</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>21..23</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>25..39</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid; up to 10 X residues may be inserted</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>41..42</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>KXXKXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXKXXK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="57">

<INSDSeq>

<INSDSeq\_length>6</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>1</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>This residue is linked N-terminally to a peptide having an N-terminal Gly-Gly and a gap of undefined length</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>5..6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GGAGXX</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="58">

<INSDSeq>

<INSDSeq\_length>45</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..45</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>CDS</INSDFeature\_key>

<INSDFeature\_location>1..45</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>translation</INSDQualifier\_name>

<INSDQualifier\_value>MTGNKPETDVYEIL</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>protein\_id</INSDQualifier\_name>

<INSDQualifier\_value>59</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>atgaccggaaataaacctgaaaccgatgtttacgaaattttatga</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="59">

<INSDSeq>

<INSDSeq\_length>14</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..14</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>MTGNKPETDVYEIL</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="60">

<INSDSeq>

<INSDSeq\_length>22</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..22</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Drosophila melanogaster</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>acattgaatctcataccacttt</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="61">

<INSDSeq>

<INSDSeq\_length>19</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..19</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Drosophila virilis</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>acatggatcccacgacttt</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="62">

<INSDSeq>

<INSDSeq\_length>22</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..22</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Drosophila simulans</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>gtatggcgtcgtatsgtagttt</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="63">

<INSDSeq>

<INSDSeq\_length>7</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X is S or T</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GGGJATX</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="64">

<INSDSeq>

<INSDSeq\_length>7</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>replace with I</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>replace with T</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GGGLATS</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="65">

<INSDSeq>

<INSDSeq\_length>7</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>X can be any amino acid</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GLPXRIC</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="66">

<INSDSeq>

<INSDSeq\_length>7</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GLPVRIC</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="67">

<INSDSeq>

<INSDSeq\_length>7</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GLPTRIC</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="68">

<INSDSeq>

<INSDSeq\_length>7</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>GLPDRIC</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="69">

<INSDSeq>

<INSDSeq\_length>36</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..36</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>cgaatgncccactacgaatgncacgaatgncccaca</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="70">

<INSDSeq>

<INSDSeq\_length>36</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..36</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>cgaatgacccactacgaatgbcacgaatgbcccaca</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="71">

<INSDSeq>

<INSDSeq\_length>36</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..36</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>cgaatgtcccactacgaatgvcacgaatgvcccaca</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="72">

<INSDSeq>

<INSDSeq\_length>36</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..36</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>cgaatggcccactacgaatghcacgaatghcccaca</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="73">

<INSDSeq>

<INSDSeq\_length>36</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..36</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>cgaatgccccactacgaatgdcacgaatgdcccaca</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="74">

<INSDSeq>

<INSDSeq\_length>400</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..400</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>CDS</INSDFeature\_key>

<INSDFeature\_location>join(1..79,142..212,272..400)</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>protein\_id</INSDQualifier\_name>

<INSDQualifier\_value>75</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>translation</INSDQualifier\_name>

<INSDQualifier\_value>MKTFAALLSAVTLALSVRAQAAVWSQCGGTPGWTGETTCVAGSVCTSLSSSYSQCVPGSATSSAPAAPSATTSGPAPTDGTCSASGAWPPLT</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>atgaagactttcgcagccttgctttccgctgtcactctcgcgctctcggtgcgcgcccaggcggctgtctggagtcaatgtaagtgccgctgcttttcattgatacgagactctacgccgagctgacgtgctaccgtataggtggcggtacaccgggttggacgggcgagaccacttgcgttgctggttcggtttgtacctccttgagctcagtgagcgactttcaatccgtcgtcattgctcctcatgtattgacgattggccttcatagtcatactctcaatgcgttccgggctccgcaacgtccagcgctccggcggccccctcagcgacaacttcaggccccgcacctacggacggaacgtgctcggccagcggggcatggccgccattgacctga</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="75">

<INSDSeq>

<INSDSeq\_length>92</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..92</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>MKTFAALLSAVTLALSVRAQAAVWSQCGGTPGWTGETTCVAGSVCTSLSSSYSQCVPGSATSSAPAAPSATTSGPAPTDGTCSASGAWPPLT</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="76">

<INSDSeq>

<INSDSeq\_length>9</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..9</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>AVLTYLRGE</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="77">

<INSDSeq>

<INSDSeq\_length>9</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..9</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>AVLTYLRGA</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="78">

<INSDSeq>

<INSDSeq\_length>9</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..9</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>AVPTYPRGE</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="79">

<INSDSeq>

<INSDSeq\_length>9</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..9</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>AVAIGYRGE</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="80">

<INSDSeq>

<INSDSeq\_length>8</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..8</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>AVLTYLGE</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="81">

<INSDSeq>

<INSDSeq\_length>32</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..32</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>N or T</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>11</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>A or T</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>12</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>K or Q</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>18</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>K or Q or N or D</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>21</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>R or G</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>22</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>K or I or N</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>24</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>L or V</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>VARIANT</INSDFeature\_key>

<INSDFeature\_location>29</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>M or L</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LEGXEQFINAXXIIRHPXYBXXTXBNDIXLIK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="82">

<INSDSeq>

<INSDSeq\_length>32</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..32</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="83">

<INSDSeq>

<INSDSeq\_length>32</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..32</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Pongo abelii</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="84">

<INSDSeq>

<INSDSeq\_length>32</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..32</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Papio anubis</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="85">

<INSDSeq>

<INSDSeq\_length>32</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..32</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Rhinopithecus roxellana</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LEGTEQFINAAKIIRHPNYNRITLDNDILLIK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="86">

<INSDSeq>

<INSDSeq\_length>32</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..32</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Pan paniscus</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LEGNEQFINAAKIIRHPKYNRITLNNDIMLIK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="87">

<INSDSeq>

<INSDSeq\_length>32</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..32</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Rhinopithecus bieti</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="88">

<INSDSeq>

<INSDSeq\_length>32</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..32</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Rhinopithecus roxellana</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="89">

<INSDSeq>

<INSDSeq\_length>7</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>join(1,7)</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>Cysteine Leucine thioester (Cys-Leu)</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>Lys side chain links the tripeptide ALE</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>CALRDKL</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="90">

<INSDSeq>

<INSDSeq\_length>5</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>5</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>Ser is linked to another sequence through an amide linkage between Ser and a side chain of a Lys in the other sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>LRDQS</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="91">

<INSDSeq>

<INSDSeq\_length>6</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>REGION</INSDFeature\_key>

<INSDFeature\_location>1..6</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>Circular peptide</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>SITE</INSDFeature\_key>

<INSDFeature\_location>4</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>NOTE</INSDQualifier\_name>

<INSDQualifier\_value>The side chain of Lys is linked via an amide linkage to another sequence</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>ALFKNG</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="92">

<INSDSeq>

<INSDSeq\_length>43</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..43</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>synthetic construct</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>other DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>misc\_feature</INSDFeature\_key>

<INSDFeature\_location>1..7</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>DNA fragment</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>misc\_feature</INSDFeature\_key>

<INSDFeature\_location>8..42</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>RNA fragment</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>misc\_feature</INSDFeature\_key>

<INSDFeature\_location>43</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>note</INSDQualifier\_name>

<INSDQualifier\_value>DNA fragment</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>agaccttcggagtctcctgttgaacagatagtcaaagtagatc</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="93">

<INSDSeq>

<INSDSeq\_length>216</INSDSeq\_length>

<INSDSeq\_moltype>DNA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>source</INSDFeature\_key>

<INSDFeature\_location>1..216</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>organism</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>mol\_type</INSDQualifier\_name>

<INSDQualifier\_value>genomic DNA</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

<INSDFeature>

<INSDFeature\_key>CDS</INSDFeature\_key>

<INSDFeature\_location>&lt;1..&gt;216</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>protein\_id</INSDQualifier\_name>

<INSDQualifier\_value>94</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>translation</INSDQualifier\_name>

<INSDQualifier\_value>HHAAECGFCPQQWQVLRGSLCICEGPAEGWFISRCWLWCGPQVQGFIFGEGKEGGGDRRAEASPQEFWECTW</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>catcacgcagcagaatgtggattttgtcctcaacaatggcaagttctacgtgggagtctgtgcatttgtgagggtccagctgaaggatggttcatatcaagatgttggttatggtgtgggcctcaagtccaaggctttatctttggagaaggcaaggaaggaggcggtgacagacgggctgaagcgagccctcaggagttttgggaatgcacttgg</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

<SequenceData sequenceIDNumber="94">

<INSDSeq>

<INSDSeq\_length>72</INSDSeq\_length>

<INSDSeq\_moltype>AA</INSDSeq\_moltype>

<INSDSeq\_division>PAT</INSDSeq\_division>

<INSDSeq\_feature-table>

<INSDFeature>

<INSDFeature\_key>SOURCE</INSDFeature\_key>

<INSDFeature\_location>1..72</INSDFeature\_location>

<INSDFeature\_quals>

<INSDQualifier>

<INSDQualifier\_name>ORGANISM</INSDQualifier\_name>

<INSDQualifier\_value>Homo sapiens</INSDQualifier\_value>

</INSDQualifier>

<INSDQualifier>

<INSDQualifier\_name>MOL\_TYPE</INSDQualifier\_name>

<INSDQualifier\_value>protein</INSDQualifier\_value>

</INSDQualifier>

</INSDFeature\_quals>

</INSDFeature>

</INSDSeq\_feature-table>

<INSDSeq\_sequence>HHAAECGFCPQQWQVLRGSLCICEGPAEGWFISRCWLWCGPQVQGFIFGEGKEGGGDRRAEASPQEFWECTW</INSDSeq\_sequence>

</INSDSeq>

</SequenceData>

</ST26SequenceListing>

[Annexe VII (ST.26 Annexe VII) suit]

1. NOTE : Si les divulgations jointes à une demande peuvent représenter les nucléotides ou les acides aminés par des symboles en lettres minuscules ou majuscules, lorsqu’il s’agit d’une séquence intégrée dans un listage des séquences, seules les lettres minuscules sont autorisées pour représenter une séquence de nucléotides (voir le paragraphe 13 de la norme ST.26) et seules les lettres majuscules le sont pour représenter une séquence d’acides aminés (voir le paragraphe 26 de la norme ST.26). [↑](#footnote-ref-2)