

Standard für die Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in Patentanmeldungen

Einführung

1. Mit dem vorliegenden Standard soll auf eine standardisierte Darstellung von Nucleotid- und Aminosäuresequenzprotokollen in Patentanmeldungen hingewirkt werden. Der Standard soll eine genauere und qualitativ bessere Darstellung von Nucleotid- und Aminosäuresequenzen sicherstellen, die Darstellung und Verbreitung der Sequenzen zum Nutzen der Anmelder, der Öffentlichkeit und der Prüfer erleichtern, die Recherche von Sequenzdaten vereinfachen und den Austausch dieser Daten in elektronischer Form sowie ihre Aufnahme in computergestützte Datenbanken ermöglichen.

Definitionen

2. Im Rahmen dieses Standards gelten folgende Definitionen:

- i) Unter einem „Sequenzprotokoll“ ist ein Teil der Beschreibung der Anmeldung in der eingereichten Fassung oder ein zur Anmeldung nachgereichtes Schriftstück zu verstehen, das die Nucleotid- und/oder Aminosäuresequenzen im Einzelnen offenbart und sonstige verfügbare Angaben enthält.
- ii) In das Protokoll dürfen nur unverzweigte Sequenzen von mindestens 4 Aminosäuren bzw. unverzweigte Sequenzen von mindestens 10 Nucleotiden aufgenommen werden. Verzweigte Sequenzen, Sequenzen mit weniger als 4 genau definierten Nucleotiden oder Aminosäuren sowie Sequenzen mit Nucleotiden oder Aminosäuren, die nicht in Anhang 2, Tabellen 1, 2, 3 und 4 aufgeführt sind, sind ausdrücklich ausgeschlossen.
- iii) Unter „Nucleotiden“ sind nur Nucleotide zu verstehen, die mittels der in Anhang 2, Tabelle 1 aufgeführten Symbole wiedergegeben werden können. Modifikationen wie z. B. methylierte Basen können nach der Anleitung in Anhang 2, Tabelle 2 beschrieben werden, sind aber in der Nucleotidsequenz nicht explizit auszuweisen.
- iv) Unter „Aminosäuren“ sind die in Anhang 2, Tabelle 3 aufgeführten gängigen L-Aminosäuren aus den natürlich vorkommenden Proteinen zu verstehen. Aminosäuresequenzen, die mindestens eine D-Aminosäure enthalten, fallen nicht unter diese Definition. Aminosäuresequenzen, die posttranslational modifizierte Aminosäuren enthalten, können mittels der in Anhang 2, Tabelle 3 aufgeführten Symbole wie die ursprünglich translatierte Aminosäure beschrieben werden, wobei die modifizierten Positionen wie z. B. Hydroxylierungen oder Glykosilierungen nach der Anleitung in Anhang 2, Tabelle 4 beschrieben werden, diese Modifikationen aber in der Aminosäuresequenz nicht explizit auszuweisen sind. Unter die vorstehende Definition fallen auch Peptide oder Proteine, die anhand der in Anhang 2, Tabelle 3 aufgeführten Symbole sowie einer an anderer Stelle aufgenommenen Beschreibung, die beispielsweise Aufschluss über ungewöhnliche Bindungen, Quervernetzungen (z. B. Disulfidbrücken) und „end caps“, Nichtpeptidbindungen usw. gibt, als Sequenz wiedergegeben werden können.
- v) Unter „Sequenzkennzahl“ ist eine Zahl zu verstehen, die jeder im Protokoll aufgeführten Sequenz als SEQ ID NO zugewiesen wird.
- vi) Die „numerische Kennzahl“ ist eine dreistellige Zahl, die für ein bestimmtes Datenelement steht.
- vii) Unter „sprachneutralem Vokabular“ ist ein festes Vokabular zu verstehen, das im Sequenzprotokoll zur Wiedergabe der vom Hersteller einer Sequenzdatenbank vorgeschriebenen wissenschaftlichen Begriffe verwendet wird (dazu gehören wissenschaftliche Namen, nähere Bestimmungen und ihre Entsprechungen im festen Vokabular, die Symbole in Anhang 2, Tabellen 1, 2, 3 und 4 und die Merkmalschlüssel in Anhang 2, Tabellen 5 und 6).

Sequenzprotokoll

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

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

5. Jeder Sequenz wird eine eigene Sequenzkennzahl zugeteilt. Die Kennzahlen beginnen mit 1 und setzen sich in aufsteigender Reihenfolge als ganze Zahlen fort. Gibt es zu einer Kennzahl keine Sequenz, so ist am Anfang der auf die SEQ ID NO folgenden Zeile unter der numerischen Kennzahl <400> der Code 000 an-

zugeben. Unter der numerischen Kennzahl <160> ist die Gesamtzahl der SEQ ID NOs anzugeben, und zwar unabhängig davon, ob im Anschluss an eine SEQ ID NO eine Sequenzkennzahl oder der Code 000 folgt.

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

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

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

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

Nucleotidsequenzen

Zu verwendende Symbole

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

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

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

Zu verwendendes Format

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

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

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

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

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

Aminosäuresequenzen

Zu verwendende Symbole

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

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

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

beschrieben wird. Diese Codes dürfen in der Beschreibung oder im Merkmalsteil des Sequenzprotokolls, nicht jedoch in der Sequenz selbst verwendet werden (s. auch Nr. 32). Das Symbol „Xaa“ steht immer nur für eine einzige unbekannte oder modifizierte Aminosäure.

Zu verwendendes Format

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

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

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

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

Sonstige verfügbare Angaben im Sequenzprotokoll

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

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

Obligatorische Datenelemente

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

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

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

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

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

<130> Bezugsnummer

27. In Sequenzprotokolle, die auf Aufforderung oder nach Vergabe eines Aktenzeichens eingereicht werden, sind neben den unter Nummer 25 genannten Datenelementen auch die folgenden aufzunehmen:

<140> Vorliegende Patentanmeldung

<141> Anmeldetag der vorliegenden Anmeldung

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

<150> Frühere Patentanmeldung

<151> Anmeldetag der früheren Anmeldung

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

<220> Merkmal

<221> Name/Schlüssel

<222> Lage

<223> Sonstige Informationen

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

<220> Merkmal

<223> Sonstige Angaben

Fakultative Datenelemente

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

Angabe von Merkmalen

32. Merkmale, die (unter der numerischen Kennzahl <220>) zu einer Sequenz angegeben werden, sind durch die in Anhang 2, Tabellen 5 und 6¹ aufgeführten „Merkmalschlüssel“ zu beschreiben.

Freier Text

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

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

35. Freier Text soll in deutscher oder englischer Sprache abgefasst sein.

36. Enthält das Sequenzprotokoll, das Bestandteil der Beschreibung ist, freien Text, so ist dieser im Hauptteil der Beschreibung in der dort verwendeten Sprache zu wiederholen. Es wird empfohlen, den in der Sprache des Hauptteils der Beschreibung abgefassten freien Text in einen besonderen Abschnitt der Beschreibung mit der Überschrift „Sequenzprotokoll - freier Text“ aufzunehmen.

Nachgereichtes Sequenzprotokoll

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

38. Sequenzprotokolle, die nicht zusammen mit der Anmeldung eingereicht worden sind, dienen nicht der Offenbarung der Erfindung.

¹ Diese Tabellen enthalten Auszüge aus den Merkmalstabellen „DDBJ/EMBL/GenBank Feature Table“ (Nucleotidsequenzen) und „SWISS PROT Feature Table“ (Aminosäuresequenzen).

Sequenzprotokoll in maschinenlesbarer Form

39. Das in der Anmeldung enthaltene Sequenzprotokoll ist zusätzlich auch in maschinenlesbarer Form einzureichen.

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

41. Das gesamte ausdrückbare Exemplar des Sequenzprotokolls muss in einer einzigen Datei enthalten sein, die nach Möglichkeit auf einer einzigen Diskette oder einem sonstigen elektronischen Datenträger aufgezeichnet sein soll (zum Beispiel optischer Datenträger wie CD). Diese Datei ist mittels der IBM²-Codetabellen (Code Page) 437, 932³ oder einer kompatiblen Codetabelle zu codieren. Eine Codetabelle, wie sie z. B. für japanische, chinesische, kyrillische, arabische, griechische oder hebräische Schriftzeichen benötigt wird, gilt als kompatibel, wenn sie das lateinische Alphabet und die arabischen Ziffern denselben Hexadezimalpositionen zuordnet wie die genannten Codetabellen.

42. Die maschinenlesbare Fassung soll mit spezieller Software wie z.B. PatentIn erstellt werden; sie kann mit beliebigen Mitteln erstellt werden, solange das Sequenzprotokoll auf einer Diskette oder einem sonstigen elektronischen Träger und in einem kompatiblen Dateiformat eingereicht wird (z.B. RTF, formatiertes TXT), das mit einem Betriebssystem für Personal Computer gelesen werden kann.

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

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

45. Wird die Diskette oder das andere elektronische Medium erst nach dem Tag der Anmeldung eingereicht, so sind auf den Etiketten auch der Anmeldetag und das Aktenzeichen anzugeben.

46. Jeder Berichtigung des geschriebenen Sequenzprotokolls, jeder Berichtigung eines offensichtlichen Fehlers im geschriebenen Sequenzprotokoll und jeder vorgenommenen Änderung, die sich auf ein geschriebenes Sequenzprotokoll bezieht, ist ein maschinenlesbares Sequenzprotokoll beizufügen, in dem diese Berichtigung oder Änderung berücksichtigt worden ist.

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

³ Die genannten Codetabellen gelten für Personal Computer praktisch als Standard.