

Anhang 1

Numerische Kennzahlen

In Sequenzprotokollen, die zu Anmeldungen eingereicht werden, dürfen nur die nachstehenden numerischen Kennzahlen verwendet werden. Die Überschriften der nachstehenden Datenelementrubriken dürfen in den Sequenzprotokollen nicht erscheinen. Die numerischen Kennzahlen der obligatorischen Datenelemente, d. h. der Datenelemente, die in alle Sequenzprotokolle aufgenommen werden müssen (s. Nr. 25 dieses Standards: Kennziffern 110, 120, 160, 210, 211, 212, 213 und 400), und die numerischen Kennzahlen der Datenelemente, die in den in diesem Standard genannten Fällen aufgenommen werden müssen (s. Nr. 26, 27, 28, 29 und 30 dieses Standards: Kennzahlen 130, 140, 141, 150 und 151 sowie 220 bis 223), sind mit dem Buchstaben „O“ gekennzeichnet. Die numerischen Kennzahlen der fakultativen Datenelemente (s. Nr. 31 dieses Standards) sind mit dem Buchstaben „F“ gekennzeichnet.

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

<212>	Art	O	ren Art des in SEQ ID NO: x sequenzierten Moleküls, und zwar entweder DNA, RNA oder PRT; enthält eine Nucleotidsequenz sowohl DNA- als auch RNA-Fragmente, so ist „DNA“ anzugeben; zusätzlich ist das kombinierte DNA-/RNA-Molekül im Merkmalsteil unter <220> bis <223> näher zu beschreiben.
<213>	Organismus	O	Gattung/Art (d. h. wissenschaftlicher Name), „künstliche Sequenz“ oder „unbekannt“
<220>	Merkmal	O in den Fällen nach Nr. 29 und 30 Standard	Freilassen; siehe Nr. 29 und 30 Standard; Beschreibung biologisch signifikanter Stellen in der Sequenz gemäß SEQ ID NO: x) (kann je nach der Zahl der angegebenen Merkmale mehrmals vorkommen).
<221>	Name/Schlüssel	O in den Fällen nach Nr. 29 Standard	Siehe Nr. 29 Standard; es dürfen nur die in Anhang 2, Tabelle 5 oder 6 beschriebenen Schlüssel verwendet werden.
<222>	Lage	O in den Fällen nach Nr. 29 Standard	Siehe Nr. 29 Standard; - von (Nummer der ersten Base/Aminosäure des Merkmals) - bis (Nummer der letzten Base/Aminosäure des Merkmals) - Basenpaare (Ziffern verweisen auf die Positionen der Basenpaare in einer Nucleotidsequenz) - Aminosäuren (Ziffern verweisen auf die Positionen der Aminosäurereste in einer Aminosäuresequenz) - Angabe, ob sich das Merkmal auf dem zum Strang des Sequenzprotokolls komplementären Strang befindet
<223>	Sonstige Angaben	O in den Fällen nach Nr. 29 und 30 Standard	Siehe Nr. 29 und 30 Standard; sonstige relevante Angaben, wobei sprachneutrales Vokabular oder freier Text (möglichst in deutscher oder englischer Sprache) zu verwenden ist; freier Text ist im Hauptteil der Beschreibung in der dort verwendeten Sprache zu wiederholen (s. Nr. 36 Standard); enthält die Sequenz eine der in Anhang 2, Tabellen 2 und 4 aufgeführten modifizierten Basen oder modifizierten/seltenen L-Aminosäuren, so ist für diese Base oder Aminosäure das dazugehörige Symbol aus Anhang 2, Tabellen 2 und 4 zu verwenden.
<300>	Veröffentlichungsangaben	F	Freilassen; dieser Abschnitt ist für jede relevante Veröffentlichung zu wiederholen.

<301>	Verfasser	F	
<302>	Titel	F	Titel der Veröffentlichung
<303>	Zeitschrift	F	Name der Zeitschrift, in der die Daten veröffentlicht wurden
<304>	Band	F	Band der Zeitschrift, in dem die Daten veröffentlicht wurden
<305>	Heft	F	Nummer des Hefts der Zeitschrift, in dem die Daten veröffentlicht wurden
<306>	Seiten	F	Seiten der Zeitschrift, auf denen die Daten veröffentlicht wurden
<307>	Datum	F	Datum der Zeitschrift, an dem die Daten veröffentlicht wurden; Angabe nach Möglichkeit entsprechend dem WIPO-Standard ST. 2 (CCYY MM DD)
<308>	Eingangsnummer in der Datenbank	F	von der Datenbank zugewiesene Eingangsnummer einschließlich Datenbankbezeichnung
<309>	Datenbank-Eingabedatum	F	Datum der Eingabe in die Datenbank; Angabe entsprechend dem WIPO-Standard ST. 2 (CCYY MM DD)
<310>	Dokumentnummer	F	Nummer des Dokuments, nur bei Patentdokumenten; die vollständige Nummer hat nacheinander Folgendes zu enthalten: den zweibuchstabigen Code entsprechend dem WIPO-Standard ST. 3, die Veröffentlichungsnummer entsprechend dem WIPO-Standard ST. 6 und den Code für die Dokumentenart nach dem WIPO-Standard ST. 16
<311>	Anmeldetag	F	Anmeldetag des Dokuments, nur bei Patentdokumenten; Angabe entsprechend WIPO-Standard ST. 2 (CCYY MM DD)
<312>	Veröffentlichungsdatum	F	Datum der Veröffentlichung des Dokuments; nur bei Patentdokumenten; Angabe entsprechend WIPO-Standard ST. 2 (CCYY MM DD)
<313>	Relevante Reste in SEQ ID NO: x von bis	F	
<400>	Sequenz	O	SEQ ID NO: x sollte in der der Sequenz vorausgehenden Zeile hinter der numerischen Kennzahl stehen (s. Anlage 3).

Symbole für Nucleotide und Aminosäuren und Merkmalstabellen

Tabelle 1: Liste der Nucleotide

Symbol	Bedeutung	Ableitung der Bezeichnung
a	a	Adenin
g	g	Guanin
c	c	Cytosin
t	t	Thymin
u	u	Uracil
r	g oder a	Purin
y	t/u oder c	Pyrimidin
m	a oder c	Amino
k	g oder t/u	Keto
s	g oder c	starke Bindungen 3 H-Brücken
w	a oder t/u	schwache (e: weak) Bindungen 2 H-Brücken
b	b oder c oder t/u	nicht a
d	a oder g oder t/u	nicht c
h	a oder c oder t/u	nicht g
v	a oder g oder c	nicht t, nicht u
n	a oder g oder c oder t/u, unbekannt oder sonstige	beliebig (e: any)

Tabelle 2: Liste der modifizierten Nucleotide

Symbol	Bedeutung
ac4c	4-Acetylcytidin
chm5u	5-(Carboxyhydroxymethyl)uridin
cm	2'-O-Methylcytidin
cmnm5s2u	5-Carboxymethylaminomethyl-2-thiouridin
cmnm5u	5-Carboxymethylaminomethyluridin
d	Dihydrouridin
fm	2'-O-Methylpseudouridin
gal q	beta, D-Galactosylqueuosin
gm	2'-O-Methylguanosin
i	Inosin
i6a	N6-Isopentenyladenosin
m1a	1-Methyladenosin
m1f	1-Methylpseudouridin
m1g	1-Methylguanosin
m1i	1-Methylinosin
m22g	2,2-Dimethylguanosin
m2a	2-Methyladenosin
m2g	2-Methylguanosin
m3c	3-Methylcytidin
m5c	5-Methylcytidin
m6a	N6-Methyladenosin
m7g	7-Methylguanosin
mam5u	5-Methylaminomethyluridin
mam5s2u	5-Methoxyaminomethyl-2-thiouridin
man q	beta, D-Mannosequeuosin

mcm5s2u	5-Methoxycarbonylmethyl-2-thiouridin
mcm5u	5-Methoxycarbonylmethyluridin
mo5u	5-Methoxyuridin
ms2i6a	2-Methylthio-N6-isopentenyladenosin
ms2t6a	N-((9-beta-D-Ribofuranosyl-2-methylthiopurin-6-yl) carbamoyl) threonin
mt6a	N-((9-beta-D-Ribofuranosylpurin-6-yl)N-methylcarbamoyl)threonin
mv	Uridin-5-oxyessigsäuremethylester
o5u	Uridin-5-oxyessigsäure (v)
osyw	Wybutoxosin
p	Pseudouridin
q	Queosin
s2c	2-Thiocytidin
s2t	5-Methyl-2-thiouridin
s2u	2-Thiouridin
s4u	4-Thiouridin
t	5-Methyluridin
t6a	N-((9-beta-D-Ribofuranosylpurin-6-yl) carbamoyl)threonin
tm	2'-O-Methyl-5-methyluridin
um	2'-O-Methyluridin
yw	Wybutosin
x	3-(3-Amino-3-carboxypropyl)uridin, (acp3)u

Tabelle 3: Liste der Aminosäuren

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

Tabelle 4: Liste der modifizierten und seltenen Aminosäuren

Symbol	Bedeutung
Aad	2-Aminoadipinsäure

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

Tabelle 5: Liste der Merkmalschlüssel zu Nucleotidsequenzen

Schlüssel	Beschreibung
allele	Ein verwandtes Individuum oder ein verwandter Stamm enthält stabile alternative Formen desselben Gens und unterscheidet sich an dieser (und vielleicht an anderer) Stelle von der vorliegenden Sequenz.
attenuator	1. Region einer DNA, in der die Beendigung der Transkription reguliert wird und die Expression einiger bakterieller Operons gesteuert wird 2. zwischen dem Promotor und dem ersten Strukturgen liegender Sequenzabschnitt, der eine partielle Beendigung der Transkription bewirkt
C_region	konstante Region der leichten und schweren Immunglobulinketten und der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; enthält je nach Kette ein oder mehrere Exons
CAAT_signal	CAAT-Box; Teil einer konservierten Sequenz, der etwa 75 Basenpaare stromaufwärts vom Startpunkt der eukaryontischen Transkriptionseinheiten liegt und an der RNA-Polymerase-Bindung beteiligt sein kann; Konsensussequenz = GG (C oder T) CAATCT
CDS	codierende Sequenz; Sequenz von Nucleotiden, die mit der Sequenz der Aminosäuren in einem Protein übereinstimmt (beinhaltet Stopcodon); Merkmal schließt eine mögliche Translation der Aminosäure ein
conflict	Unabhängige Determinierungen „derselben“ Sequenz unterscheiden sich an dieser Stelle oder in dieser Region voneinander.
D-loop	D-Schleife; Region innerhalb der mitochondrialen DNA, in der sich ein kürzeres RNA-Stück mit einem Strang der doppelsträngigen DNA paart und dabei den ursprünglichen Schwesterstrang in dieser Region verdrängt; dient auch zur Beschreibung der Verdrängung einer einzelsträngigen Region des DNA-Doppelstrangs durch einen einzelsträngigen Eindringling bei der durch ein recA-Protein ausgelösten Reaktion
D-segment	Diversity-Region der schweren Kette von Immunglobulin und der Beta-Kette eines T-Zell-Rezeptors

enhancer	eine als Cis-Element wirkende Sequenz, die die Aktivität (einiger) eukaryontischer Promotoren verstärkt und in beliebiger Richtung und Position zum Promotor (stromaufwärts oder -abwärts) funktioniert
exon	Region des Genoms, die für einen Teil der gespleißten mRNA codiert; kann 5'UTR, alle CDSs und 3'UTR enthalten
GC_signal	GC-Box; eine konservierte, GC-reiche Region stromaufwärts vom Startpunkt der eukaryontischen Transkriptionseinheiten, die in mehreren Kopien und in beiden Richtungen vorkommen kann; Konsensussequenz = GGGCGG
gene	biologisch signifikante Region, die als Gen bezeichnet wird und einen Namen trägt
iDNA	intervenierende DNA; DNA, die durch verschiedene Arten der Rekombination eliminiert wird
intron	DNA-Abschnitt, der transkribiert, aber beim Zusammenspleißen der ihn umrahmenden Sequenzen (Exons) aus dem Transkript wieder herausgeschnitten wird
J_segment	J-Kette (Verbindungskette) zwischen den leichten und den schweren Immunglobulinketten und den Alpha-, Beta- und Gamma-Ketten der T-Zell-Rezeptoren
LTR	lange Terminalwiederholung, eine an den beiden Enden einer gegebenen Sequenz direkt wiederholte Sequenz, wie sie für Retroviren typisch ist
mat_peptide	für ein reifes Peptid oder Protein codierende Sequenz; Sequenz, die im Anschluss an eine posttranslationale Modifizierung für das reife oder endgültige Peptid- oder Proteinprodukt codiert; schließt im Gegensatz zur entsprechenden CDS das Stopcodon nicht ein
misc_binding	Stelle in einer Nucleinsäure, die einen anderen Teil, der nicht durch einen anderen Bindungsschlüssel (primer_bind oder protein_bind) beschrieben werden kann, kovalent oder nicht kovalent bindet
misc_difference	Die Merkmalsequenz unterscheidet sich von der im Eintrag und kann nicht durch einen anderen Unterscheidungsschlüssel (conflict, unsure, old_sequence, mutation, variation, allele bzw. modified_base) beschrieben werden.
misc_feature	biologisch signifikante Region, die nicht durch einen anderen Merkmalschlüssel beschrieben werden kann; neues oder seltenes Merkmal
misc_recomb	Stelle, an der ein allgemeiner, ortsspezifischer oder replikativer Rekombinationsvorgang stattfindet, bei dem die DNA-Doppelhelix aufgebrochen und wieder zusammengefügt wird, und die nicht durch andere Rekombinationsschlüssel (iDNA oder Virion) oder den betreffenden Herkunftsschlüssel (/insertions_seq, /transposon, /proviral) beschrieben werden kann
misc_RNA	Transkript oder RNA-Produkt, das nicht durch andere RNA-Schlüssel (prim_transcript, precursor_RNA, mRNA, 5'clip, 3'clip, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, rRNA, tRNA, scRNA oder snRNA) beschrieben werden kann
misc_signal	Region, die ein Signal enthält, das die Genfunktion oder -expression steuert oder ändert, und die nicht durch andere Signalschlüssel (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator oder rep_origin) beschrieben werden kann
misc_structure	Sekundär-, Tertiär- oder sonstige Struktur oder Konformation, die nicht durch andere Strukturschlüssel (stem_loop oder D-loop) beschrieben werden kann
modified_base	Das angegebene Nucleotid ist ein modifiziertes Nucleotid und ist durch das angegebene Molekül (ausgedrückt durch die modifizierte Base) zu ersetzen.
mRNA	messenger-RNA (Boten-RNA); enthält eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon) und eine 3'-nichttranslatierte Region (3'UTR)
mutation	Ein verwandter Stamm weist an dieser Stelle eine plötzliche, erbliche Sequenzveränderung auf.
N_region	zusätzliche Nucleotide, die zwischen neu geordnete Immunglobulinabschnitte eingefügt werden
old_sequence	Die vorliegende Sequenz stellt eine geänderte Version der früher an dieser Stelle befindlichen Sequenz dar.
polyA_signal	Erkennungsregion, die zur Endonuclease-Spaltung eines RNA-Transkripts mit anschließender Polyadenylierung nötig ist; Konsensussequenz: AATAAA
polyA_site	Stelle auf einem RNA-Transkript, an der durch posttranslationale Polyadenylierung Adeninreste eingefügt werden
precursor_RNA	noch nicht gereifte RNA-Spezies; kann eine am 5'-Ende abzuschneidende Region (5'Clip), eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron), eine 3'-nichttranslatierte Region (3'UTR) und eine am 3'-Ende abzuschneidende Region (3'Clip) enthalten

prim_transcript	primäres (ursprüngliches, nicht prozessiertes) Transkript; enthält eine am 5'-Ende abzuschneidende Region (5'Clip), eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron), eine 3'-nichttranslatierte Region (3'UTR) und eine am 3'-Ende abzuschneidende Region (3'Clip)
primer_bind	nichtkovalente Primer-Bindungsstelle für die Initiierung der Replikation, Transkription oder reversen Transkription; enthält eine oder mehrere Stellen für synthetische Elemente, z. B. PCR-Primerelemente
promoter	Region auf einem DNA-Molekül, die an der Bindung der RNA-Polymerase beteiligt ist, die die Transkription initiiert
protein_bind	nichtkovalente Protein-Bindungsstelle auf der Nucleinsäure
RBS	ribosomale Bindungsstelle
repeat_region	Genomregion mit repetitiven Einheiten
repeat_unit	einzelnes Repeat (repetitive Einheit)
rep_origin	Replikationsursprung; Startpunkt der Duplikation der Nucleinsäure, durch die zwei identische Kopien entstehen
rRNA	reife ribosomale RNA; RNA-Komplex des Ribonucleoprotein-Partikels (Ribosom), der Aminosäuren zu Proteinen zusammenfügt
S_region	Switch-Region der schweren Immunglobulinketten; beteiligt am Umbau schwerer DNA-Ketten, der zur Expression einer anderen Immunglobulin-Klasse aus derselben B-Zelle führt
satellite	viele tandemartig hintereinandergeschaltete (identische oder verwandte) Repeats einer kurzen basischen repetitiven Einheit; viele davon unterscheiden sich in der Basenzusammensetzung oder einer anderen Eigenschaft vom Genomdurchschnitt und können so von der Hauptmasse der genomischen DNA abgetrennt werden
scRNA	kleine cytoplasmische RNA; eines von mehreren kleinen cytoplasmischen RNA-Molekülen im Cytoplasma und (manchmal) im Zellkern eines Eukaryonten
sig_peptide	für ein Signalpeptid codierende Sequenz; Sequenz, die für eine N-terminale Domäne eines sekretorischen Proteins codiert; diese Domäne spielt bei der Anheftung des naszierenden Polypeptids an die Membran eine Rolle; Leader-Sequenz
snRNA	kleine Kern-RNA; eine der vielen kleinen RNA-Formen, die nur im Zellkern vorkommen; einige der snRNAs spielen beim Spleißen oder bei anderen RNA-verarbeitenden Reaktionen eine Rolle
source	Bezeichnet die biologische Herkunft des genannten Sequenzabschnitts; die Angabe dieses Schlüssels ist obligatorisch; jeder Eintrag muss mindestens einen Herkunftsschlüssel aufweisen, der die gesamte Sequenz umfaßt; es dürfen zu jeder Sequenz auch mehrere Herkunftsschlüssel angegeben werden.
stem_loop	Haarnadelschleife; eine Doppelhelix-Region, die durch Basenpaarung zwischen benachbarten (invertierten) komplementären Sequenzen in einem RNA- oder DNA-Einzelstrang entsteht
STS	Sequence Tagged Site; kurze, nur als Einzelkopie vorkommende DNA-Sequenz, die einen Kartierungspunkt auf dem Genom bezeichnet und durch PCR ermittelt werden kann; eine Region auf dem Genom kann durch Bestimmung der Reihenfolge der STSs kartiert werden
TATA_signal	TATA-Box; Goldberg-Hogness-Box; ein konserviertes AT-reiches Septamer, das sich rund 25 Basenpaare vor dem Startpunkt jeder eukaryontischen RNA-Polymerase-II-Transkriptionseinheit befindet und bei der Positionierung des Enzyms für eine korrekte Initiation eine Rolle spielen kann; Konsensussequenz = TATA(A oder T)A(A oder T)
terminator	DNA-Sequenz, die entweder am Ende des Transkripts oder neben einer Promotor-Region liegt und bewirkt, dass die RNA-Polymerase die Transkription beendet; kann auch die Bindungsstelle eines Repressor-Proteins sein
transit_peptide	für Transitpeptid codierende Sequenz; Sequenz, die für eine N-terminale Domäne eines im Zellkern codierten Organellen-Proteins codiert; diese Domäne ist an der posttranslationalen Einschleusung des Proteins in die Organelle beteiligt
tRNA	reife transfer-RNA, ein kleines RNA-Molekül (75-85 Basen lang), das die Translation einer Nucleinsäure-Sequenz in eine Aminosäure-Sequenz vermittelt
unsure	Der Autor kennt die Sequenz in dieser Region nicht genau.
V_region	variable Region der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; codiert für das variable Aminoende; kann aus V-, D-, N- und J-Abschnitten bestehen
V-segment	variabler Abschnitt der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; codiert für den Großteil der variablen Region(V_region) und die letzten Aminosäuren des Leader-Peptids
variation	Ein verwandter Stamm enthält stabile Mutationen desselben Gens (z. B. RFLPs, Poly-

	morphismen usw.), die sich an dieser (und möglicherweise auch an anderer) Stelle von der vorliegenden Sequenz unterscheiden.
3'clip 3'UTR	3'-äußerste Region eines Precursor-Transkripts, die bei der Prozessierung abgeschnitten wird Region am 3'-Ende eines reifen Transkripts (nach dem Stopcodon), die nicht in ein Protein translatiert wird
5'clip 5'UTR	5'-äußerste Region eines Precursor-Transkripts, die bei der Prozessierung abgeschnitten wird Region am 5'-Ende eines reifen Transkripts (vor dem Initiationscodon), die nicht in ein Protein translatiert wird
-10_signal	Pribnow-Box; konservierte Region rund 10 Basenpaare stromaufwärts vom Startpunkt der bakteriellen Transkriptionseinheiten, die bei der Bindung der RNA-Polymerase eine Rolle spielt; Konsensussequenz = TAtAaT
-35_signal	konserviertes Hexamer rund 35 Basenpaare stromaufwärts vom Startpunkt der bakteriellen Transkriptionseinheiten; Konsensussequenz = TTGACa [] oder TGTTGACA []

Tabelle 6: Liste der Merkmalschlüssel zu Proteinsequenzen

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

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

Muster eines Sequenzprotokolls

<110> Smith, John; Smithgene Inc.

<120> Beispiel für ein Sequenzprotokoll

<130> 01-00001

<140> PCT/EP98/00001

<141> 1998-12-31

<150> US 08/999 999

<151> 1997-10-15

<160> 4

<170> PatentIn Version 2.0

<210> 1

<211> 389

<212> DNA

<213> Paramecium sp.

<220>

<221> CDS

<222> (279) ... (389)

<300>

<301> Doe, Richard

<302> Isolation and Characterization of a Gene Encoding a Protease
from Paramecium sp.

<303> Journal of Genes

<304> 1

<305> 4

<306> 1-7

<307> 1988-06-31

<308> 123456

<309> 1988-06-31

<400> 1

agctgtagtc	attcctgtgt	cctctctct	ctgggcttct	caccctgcta	atcagatctc	60
agggagagtg	tcttgacct	cctctgcctt	tgccagctca	caggcaggca	ggcaggcagc	120
tgatgtggca	attgctggca	gtgccacagg	ctttcagcc	aggcttaggg	tgggtccgc	180
cgcgcgcgcg	cggcccctct	cgcgctcctc	tcgcgctct	ctctcgctct	cctctcgctc	240
ggacctgatt	aggtgagcag	gaggaggggg	cagttagc	atg gtt tca	atg ttc agc	296
				Met Val Ser	Met Phe Ser	
				1	5	
ttg tct	ttc aaa tgg	cct gga ttt	tgt ttg	ttt gtt	tgt ttg ttc	caa
Leu Ser	Phe Lys Trp	Pro Gly Phe	Cys Leu	Phe Val	Cys Leu Phe	Gln

