

F1-Praktikum Genetik

Protokoll Versuch 4: Nutzung des Internet in der Biologie

Gruppe 8, Manfred Depner & Susanne Duncker

Einleitung

Das Internet ist ein äußerst effizientes Hilfsmittel in der Biologie, da es z.B. in der Genetik viele DNA-Sequenzen u.ä. in einer rund um die Uhr verfügbaren und (meist) für jedermann zugänglichen Datenbank bietet. Um uns mit der Nutzung dieser Möglichkeiten vertraut zu machen, bearbeiteten wir die folgenden Aufgaben.

Aufgabe 1: Suchen Sie die cDNA-Sequenz des menschlichen Insulins, machen Sie einen Ausdruck und markieren Sie das Start- und das Stopp-Codon

Folgender Weg wurde zum Ergebnis verwendet:

- <http://www.ncbi.nlm.nih.gov/> → links: Genomic biology, Unterpunkt „The human Genom“
 - o rechts: Human
 - o Search Genes: from Homo Sapiens, with words insulin
 - o Ergebnis 15: INS, Insulin (Human), rechts: MGC cDNA clone
 - o Eintrag 1 ([BC005255](#)), auswählen, ganz unten ist die Sequenz
- im entsprechenden ausgedruckten Dokument (Anhang 1) wurden das Start- und Stoppcodon gelb hinterlegt

Aufgabe 2: Übersetzen Sie die cDNA-Sequenz in alle drei möglichen Leserahmen – wie viele Stopp-Codons finden Sie in Rahmen 1, 2 und 3?

- <http://workbench.sdsc.edu> → Account erstellen, anmelden
 - o Auswahl von „Nucleic Tools“
 - o „Add New Nucleic Sequence“
 - bei „Label“ benannten wir die Sequenz „insulin homo sapiens“
 - in „Sequence“ gaben wir die oben gefundene cDNA-Sequenz ein
 - mit einem Klick auf „Save“ war die Sequenz gespeichert und verfügbar
 - o wieder auf <http://seqtool.sdsc.edu/CGI/BW.cgi#!> → „SIXFRAME“, unten „User Entered insulin homo sapiens“ anhängen, „Run“
 - alle Einstellungen wurden von uns so gelassen, wie sie angeboten wurden, mit Ausnahme der „Frame to translate“ – hier wählten wir „3 Forward Frames“
 - „Submit“ → alle drei Leserahmen werden ausgegeben, dazu jeweils die Anzahl der Stopcodons (im Ausdruck (Anhang 2) mit einem farbigen Stern markiert)
 - Frame 1: 4 Stopp-Codons
 - Frame 2: 1 Stopp-Codon
 - Frame 3: 1 Stopp-Codon
 - longest open reading frame: Frame 3

Aufgabe 3: Vergleichen Sie die Proteinsequenz des menschlichen Insulins mit der des Schweineinsulins mit Hilfe von BLAST – warum macht es mehr Sinn, nur die Aminosäuresequenzen zu vergleichen? Warum ist es in der Anwendung besser, menschliches Insulin zur Verfügung zu haben?

- <http://www.ncbi.nlm.nih.gov/> → links auf „Genomic biology“
 - o rechts auf „Pig“

- o rechts auf GenBank
- o oben „Search Gene for Sus scrofa[orgn] Insulin“
- o Ergebnis 17: [INS2](#), preproinsulin anklicken
 - Related Sequences: Nucleotide mRNA [AF064555](#) anklicken, ganz unten ist die Sequenz
- die Sequenz wie bei Aufgabe 2 in die workbench einfügen
- sowohl den longest ORF von Schweineinsulin als auch den des menschlichen Insulins erzeugen und mit „Import Sequences“ in den Proteinbearbeitungsbereich der workbench importieren
- in der workbench die „Protein Tools“ auswählen, Schweine- und Menscheninsulinsequenzen unten anhaken und „BL2SEQ“ auswählen → „Run“
- bei Matrix Selection wählen wir „BLOSUM45, 15, 2, 0.90 RECOMMENDED“ → „Submit“
- der Ausdruck des Ergebnisses findet sich als Anhang 3 am Ende des Protokolls

Es macht mehr Sinn, die Aminosäuresequenzen anstatt der Nukleotidsequenzen zu vergleichen, da es sich bei der DNA um einen degenerierten Code handelt, in dem verschiedene Triplets für gleiche Aminosäuren codieren können → all diese verschiedenen Triplets würden als Unterschiede ausgegeben werden.

Aufgabe 4: Erstellen Sie eine Restriktionskarte der menschlichen Insulin-cDNA-Sequenz, beginnend mit ATG und endend mit dem Stopp-Codon

- <http://workbench.sdsc.edu/> → Nucleic Tools
- „User Entered insulin homo sapiens“ anhaken, TACG auswählen, „Run“
- voreingestellte Einstellungen übernehmen, nur in „Output Parameters“ „Display Translation“ auf „3 forward Frames“ einstellen sowie „Smallest Fragment Cutoff Size“ auf „10“, durch „Submit“ starten
- man erhält die Schnittstellen der verschiedenen Restriktionsenzyme sowie die einzelnen Fragmentgrößen auf einem virtuellen Gel

Aufgabe 5: Finden Sie die Restriktionskarte des Vektors pET21(+)

- <http://www.novagen.com>; Auswahl des Landes auf Deutschland, dann „Submit“
- in „Search“ pET21 in „All“ eingeben
- Auswahl von [TB063: pET-21\(+\) Vector Map](#) unter „Technical Literature“
- dann auf Download → die Restriktionskarte findet sich als Anhang 4 hinter dem Protokoll

Aufgabe 6: Entwerfen Sie Primersequenzen, mit denen Sie die menschliche Insulin-cDNA-Sequenz in den Vektor pET-21(+) inklonieren können, so dass Sie das daraus gebildete Produkt über einen C-terminalen 6xHis-Tag aufreinigen können

Primer 1: 5' GACAGGCTGCATCAGAAGAG 3'

Primer 2: 5' TCTCTCGGTGCAGGAGGCGG 3'

Aufgabe 7: Vergleichen Sie die Aminosäuresequenz menschlichen Insulins mit der des Mäuseinsulins – was fällt auf?

- auf <http://www.ncbi.nlm.nih.gov/> nach „Insulin“ in „Gene“ suchen
 - o oben die Registerkarte „Limits“ auswählen
 - o unter „Limit by Taxonomy“ „Mus musculus“ anhaken
 - o unter „Include only“ „Genomic“ anhaken
 - o unter „Exclude“ „Plasmids“ anhaken
 - o oben wieder auf „Go“

- o Treffer 2: [Ins1](#)
 - dort auf **Source Sequence** [AK007482](#)
 - dort ist ganz unten die Sequenz
- Sequenz in der workbench in Proteinsequenz umwandeln und mit menschlichem Insulin vergleichen (s.o.)

Auffällig ist, dass Mäuseinsulin nur zu 77 % mit menschlichem Insulin übereinstimmt, während Schweineinsulin zu 85 % unserem Insulin ähnelt – das Schweineinsulin ist unserem also ähnlicher als das Mäuseinsulin.

NCBI

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Nucleotide for

Limits Preview/Index History Clipboard Details

GenBank all to file

Range: from to ☐ Reverse complemented strand Features: ☐ SNP

☐ CDD ☒ MGC ☐ HPRD

☐ 1: [BC005255](#). Reports Homo sapiens insu...[gi:13528923]

[Links](#)

LOCUS BC005255 495 bp mRNA linear PRI 29-JUN-2004
DEFINITION Homo sapiens insulin, mRNA (cDNA clone MGC:12292 IMAGE:3950204), complete cds.
ACCESSION BC005255
VERSION BC005255.1 GI:13528923
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 495)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED [12477932](#)
REFERENCE 2 (bases 1 to 495)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 16 Row: m Column: 8
This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 4557670.

FEATURES
 source
 Location/Qualifiers
 1..495
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:12292 IMAGE:3950204"
 /tissue_type="Pancreas"
 /clone_lib="NIH_MGC_78"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"
 gene
 1..495
 /gene="INS"
 /db_xref="GeneID:3630"
 /db_xref="MIM:176730"
 CDS
 60..392
 /gene="INS"
 /codon_start=1
 /product="proinsulin, precursor"
 /protein_id="AAH05255.1"
 /db_xref="GI:13528924"
 /db_xref="GeneID:3630"
 /db_xref="MIM:176730"
 /translation="MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN"
 misc_difference
 414
 /gene="INS"
 /note="'C' in cDNA is 'A' in the human genome. The chimpanzee genome agrees with the cDNA sequence, suggesting that this difference is unlikely to be due to an artifact."
 misc_difference
 464..495
 /gene="INS"
 /note="polyA tail: 32 bases do not align to the human genome."
ORIGIN
 1 agccctccag gacaggctgc atcagaagag gccatcaagc agatcactgt ctttctgcca
 61 tggccctgtg gatgcgcctc ctgcccctgc tggcgctgct ggccctctgg ggacctgacc
 121 cagccgcagc ctttgtgaac caacacctgt gcggctcaca cctggtggaa gctctctacc
 181 tagtgtgcgg ggaacgaggc ttcttttaca caccgaagac ccgccgggag gcagaggacc
 241 tgcaggtggg gcaggtggag ctgggcgggg gccctggtgc aggcagcctg cagcccttgg
 301 ccctggaggg gtccctgcag aagcgtggca ttgtggaaca atgctgtacc agcatctgct
 361 ccctctacca gctggagaac tactgcaact agacgcagcc cgcaggcagc cccccaccg
 421 ccgcctcctg caccgagaga gatggaataa agcccttgaa ccaacaaaaa aaaaaaaaaa
 481 aaaaaaaaaa aaaaaa

//

[Disclaimer](#) | [Write to the Help Desk](#)
[NCBI](#) | [NLM](#) | [NIH](#)

Feb 9 2005 14:31:10



Version 3.2

SIXFRAME

Generate & Import 6 Frame Translations on a NS

Selected Sequence(s)

- insulin homo sapiens

Import Sequence(s)

Return

Help

Report Bugs

Translation table used:

The Standard Code

Note:

Stop codons (*) should be edited out of imported protein sequences before use in further analysis, as they are not recognized by most analysis programs.

Unresolved triplets are marked with a "X" character.

Frame 1, 4 stop codons

☐ insulin homo sapiens Translated - Frame 1

```
>insulin_homo_sapiens Translated - Frame 1
SPPGQAASEEAIKQITVLLPWPCGCASCPCWRCWPSGDLTQPQL*TNCAHTWWKLST*CAGNEASST
HPRPAGRQRTCRWGRWSWAGALVQAACSPWPWRGPCRSVALWNNAPASPTSWRTTATRRSPQAAPHP
PPPAPREME*SP*TNKKKKKKKKKK
```

```

1  S P P G Q A A S E E A I K Q I T V L L P      60
   agccctccaggacaggctgcacagaaggccatcaagcagatcactgtccttctgcc
61  W P C G C A S C P C W R C W P S G D L T      120
   tggccctgtggatgcgcctcctgccctgctggcgctgctggccctctggggac
   Q P Q P L * T N T C A A H T W W K L S T
121  cagccgcagccttgtgaaccaacacctgtgcggctcacacctggtggaagctcttacc  180
   * C A G N E A S S T H P R P A G R Q R T
181  tagtgtgcggggaacgaggtcttcttacacaccaagaccgcccgggaggcagaggacc  240
   C R W G R W S W A G A L V Q A A C S P W
241  tgcaggtggggcaggtggagctggcgggggccctggtgcaggcagcctgcagccctgg  300
   P W R G P C R S V A L W N N A V P A S A
301  ccctggaggggtccctgcagaagcgtggcattgtggaacaatgctgtaccagcatctgct  360
   P S T S W R T T A T R R S P Q A A P H P
361  ccctctaccagctggagaactgaactagacgcagcccgaggcagccccccaccg    420
   P P P A P R E M E * S P * T N K K K K K
```

```
421 cgcctcctgcaccgagagagatggaataaagcccttgaaccaacaaaaaaaaaaaaaa 480
    K K K K K
481 aaaaaaaaaaaaaa 495
```

Frame 2, 1 stop codon

☐ insulin homo sapiens Translated - Frame 2

```
>insulin_homo_sapiens Translated - Frame 2
ALQDR LHQKRPSSRSLSFCHGPVDAPPAPAGAAGPLGT*PSRSLCEPTPVRLTPGGSSLPSVRGTRLLH
TQDPPGGRGPAGGAGGAGRPWCRQPAALGPGVPAEAWHCGTMLYQHLLPLPAGELLQLDAARRQPPT
RLLHRERWNKALEPTKKKKKKKK
```

```

  A L Q D R L H Q K R P S S R S L S F C H
2   gccctccaggacaggctgcatcagaagaggccatcaagcagatcactgtccttctgcat 61
    G P V D A P P A P A G A A G P L G T * P
62   ggccctgtggatgcgcctcctgcccctgctggcgctgctggccctctggggacctgaccc 121
    S R S L C E P T P V R L T P G G S S L P
122  agccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacct 181
    S V R G T R L L L H T Q D P P G G R G P
182  agtgtgcggggaacgaggcttcttctacacaccaagacccgcccgggaggcagaggacct 241
    A G G A G G A G R G P W C R Q P A A L G
242  gcaggtggggcaggtggagctgggcccctggtgcaggcagcctgcagcccttggc 301
    P G G V P A E A W H C G T M L Y Q H L L
302  cctggaggggtccctgcagaagcgtggcattgtggaacaatgctgtaccagcatctgctc 361
    P L P A G E L L Q L D A A R R Q P P T R
362  cctctaccagctggagaactactgcaactagacgagcccgaggcagccccccaccgc 421
    R L L H R E R W N K A L E P T K K K K K
422  cgcctcctgcaccgagagagatggaataaagcccttgaaccaacaaaaaaaaaaaaaa 481
    K K K K K
482  aaaaaaaaaaaaaa 495
```

Frame 3, 1 stop codon

☐ insulin homo sapiens Translated - Frame 3

```
>insulin_homo_sapiens Translated - Frame 3
PSRTGCIIRRGHQADHCP SAMALWMRLPLLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYT
PKTRREAEDLQVGQVELGGGPGAGLSQPLALEGLSLQKRGIVEQCCTSI CSLYQLENYCN*TQPAGSPPPA
ASCTERDGIKPLNQKKKKKKKK
```

```

  P S R T G C I R R G H Q A D H C P S A M
3   ccctccaggacaggctgcatcagaagaggccatcaagcagatcactgtccttctgcatg 62
    A L W M R L L P L L A L L A L W G P D P
63   gccctgtggatgcgcctcctgcccctgctggcgctgctggccctctggggacctgaccca 122
    A A A F V N Q H L C G S H L V E A L Y L
123  gccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacct 182
    V C G G E R G G F F Y T P K T R R E A E D L
183  gtgtgcggggaacgaggcttcttctacacaccaagacccgcccgggaggcagaggacctg 242
    Q V G Q V E L G G G P G A G S L Q P L A
243  caggtggggcaggtggagctgggcccctggtgcaggcagcctgcagcccttggc 302
```

```
      L E G S L Q K R G I V E Q C C T S I C S
303  ctggaggggtccctgcagaagcgtggcattgtggaacaatgctgtaccagcatctgctcc 362
      L Y Q L E N Y C N * T Q P A G S P P P A
363  ctctaccagctggagaactactgcaactagacgcagcccgcaggcagccccccaccgcc 422
      A S C T E R D G I K P L N Q Q K K K K K
423  gcctcctgcaccgagagagatggaataaagcccttgaaccaacaaaaaaaaaaaaaaaa 482
      K K K K
483  aaaaaaaaaaaaaa 495
```

Frame 3 [Longest ORF], 0 stop codons

☐ insulin_homo_sapiens Translated - Longest ORF [Frame 3]

```
>insulin_homo_sapiens Translated - Longest ORF [Frame 3]
MALWMRLLPLLALLLWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGG
GPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
```

[Import Sequence\(s\)](#)[Return](#)[Help](#)[Report Bugs](#)

Copyright (C) 1999, Board of Trustees of the University of Illinois.





Version 3.2

BL2SEQ

Compare proteins to each other with BLAST

Selected Sequence(s)

- preproinsulin sus scrofa Translated - Longest ORF [Frame 3]
- insulin homo sapiens Translated - Longest ORF [Frame 3]

Return

Help

Report Bugs

preproinsulin sus scrofa Translated - Longest ORF [Frame 3] (User Entered,SIXFRAME Translated)

- and -

insulin homo sapiens Translated - Longest ORF [Frame 3] (User Entered,SIXFRAME Translated)

Query: >preproinsulin_sus_scrofa Translated - Longest ORF [Frame 3]
Length = 108

Reference: Query= preproinsulin_sus_scrofa Translated - Longest ORF [Frame 3]
(108 letters)

>insulin_homo_sapiens Translated - Longest ORF [Frame 3]
Length = 110

Score = 178 bits (593), Expect = 2e-50

Identities = 94/110 (85%), Positives = 95/110 (86%), Gaps = 2/110 (1%)

Query: 1 MALWTRLLPLLALLALWAPAPAQAFVNQHLGSHLVEALYLVCGERGFFYTPKARREAEN 60
MALW RLLPLLALLALW P PA AFVNQHLGSHLVEALYLVCGERGFFYTPK RREAE+
Sbjct: 1 MALWMRLLPLLALLALWGPDPAAAFVNQHLGSHLVEALYLVCGERGFFYTPKTRREAED 60

Query: 61 PQAGAVELGG--GLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN 108
Q G VELGG G G LQ LALEG QKRGIVEQCCTSICSLYQLENYCN
Sbjct: 61 LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN 110

Lambda	K	H
0.231	0.0972	0.289

Gapped		
Lambda	K	H
0.203	0.0410	0.120

Matrix: BLOSUM45

Gap Penalties: Existence: 15, Extension: 2

Number of Hits to DB: 143

Number of Sequences: 0

Number of extensions: 6

Number of successful extensions: 3

Number of sequences better than 10.0: 1

Number of HSP's better than 10.0 without gapping: 1

Number of HSP's successfully gapped in prelim test: 0

Number of HSP's that attempted gapping in prelim test: 0

Number of HSP's gapped (non-prelim): 1
length of query: 108
length of database: 110
effective HSP length: 17
effective length of query: 91
effective length of database: 93
effective search space: 8463
effective search space used: 8463
T: 14
A: 60
X1: 18 (6.0 bits)
X2: 51 (14.9 bits)
X3: 85 (24.9 bits)
S1: 18 (9.4 bits)
S2: 18 (9.9 bits)

[Return](#)[Help](#)[Report Bugs](#)

Citation

Algorithm Citation:

Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Program Citation:

National Center for Biotechnology Information, Bldg 38A, NIH, 8600 Rockville Pike, Bethesda, MD 20894

Copyright (C) 1999, Board of Trustees of the University of Illinois.



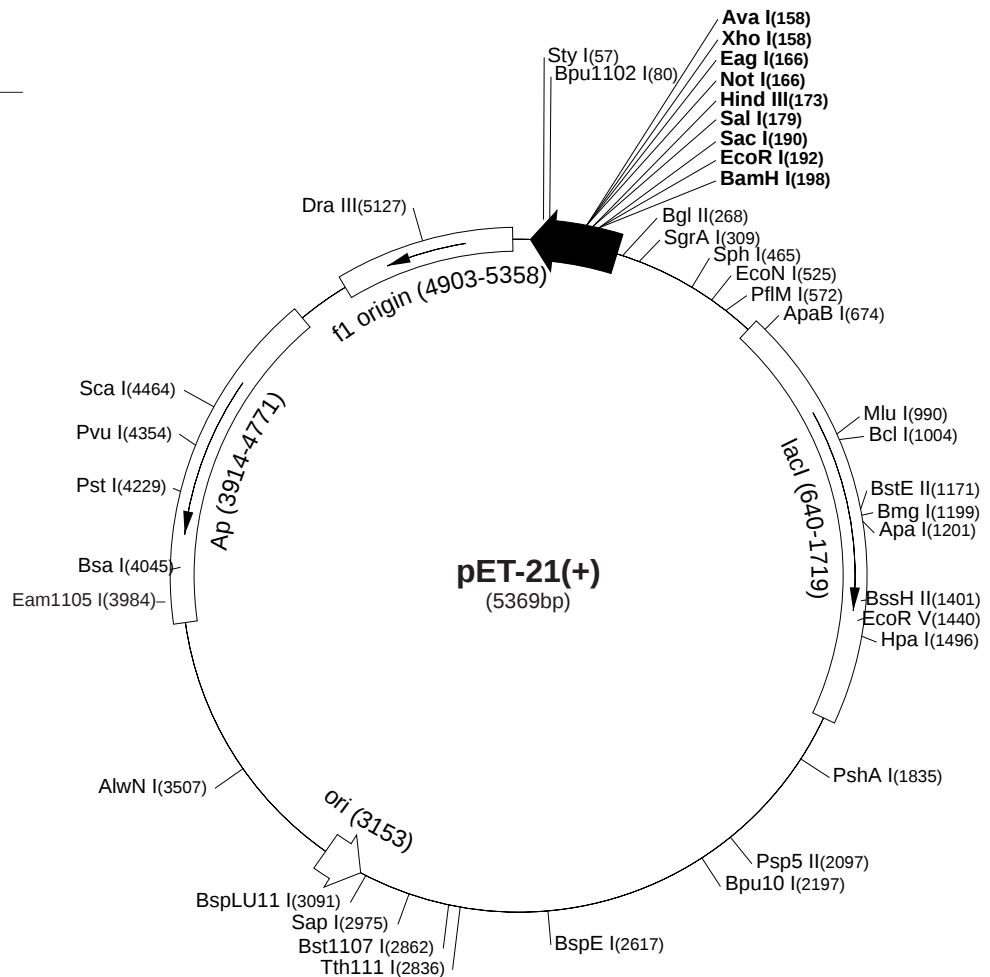
pET-21(+) Vector

TB063 12/98

pET-21(+) (Cat. No. 69770-3) is a transcription vector designed for expression from bacterial translation signals carried within a cloned insert. It therefore lacks the ribosome binding site and ATG start codon present on the pET translation vectors. A C-terminal His•Tag® sequence is available. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The f1 origin is oriented so that infection with helper phage will produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3).

pET-21(+) sequence landmarks

T7 promoter	237-253
T7 transcription start	236
Multiple cloning sites (<i>Bam</i> H I - <i>Xho</i> I)	158-203
His•Tag® coding sequence	140-157
T7 terminator	26-72
<i>lacI</i> coding sequence	640-1719
pBR322 origin	3153
<i>bla</i> coding sequence	3914-4771
f1 origin	4903-5358



pET-21(+) cloning/expression region

pET-21(+) Restriction Sites

TB063 12/98

Enzyme	# Sites	Locations					Enzyme	# Sites	Locations					Enzyme	# Sites	Locations				
AccI	2	180	2861				Bst1107I	1	2862				Psp1406I	6	652	2020	2416	4210	4583	
AccIII	8	757	1485	1816	2600	2741	BstEII	1	1171						4912					
		3043	4283	4967			BstXI	3	792	921	1044		PstI	1	4229					
AcI	80						BstYI	12					PvuI	1	4354					
AflIII	2	990	3091				Cac8I	38					PvuII	3	1590	1683	2682			
AluI	25						CjeI	24					RcaI	3	388	3811	4819			
AlwI	16						CjePI	18					RsaI	3	1137	2897	4464			
Alw21I	9	159	190	490	974	2085	CviJI	85					SacI	1	190					
		2909	3409	4570	4655		CviRI	24					Sall	1	179					
Alw44I	4	970	2905	3405	4651		DdeI	11					SapI	1	2975					
AlwNI	1	3507					DpnI	27					Sau96I	18						
Apal	1	1201					DraI	3	3850	3869	4561		Sau3AI	27						
ApaBI	1	674					DraIII	1	5127				Scal	1	4464					
ApoI	4	192	1265	4929	4940		DrdI	3	2784	3199	5082		ScrFI	19						
Aval	1	158					DrdII	2	713	5132			SfaNI	20						
Avall	7	1542	1918	2006	2097	2376	Dsal	2	427	2063			Sfcl	5	236	3356	3547	4225 5346		
		4122	4344				EaeI	5	166	298	430 1664 4372		SgrAI	1	309					
BamHI	1	198					EagI	1	166				SphI	1	465					
BanI	9	312	333	447	910	1629	Eam1105I	1	3984				Sspl	2	4788	4919				
		1759	1885	3932	5164		EarI	3	608	2975	4779		StyI	1	57					
BanII	5	190	374	388	1201	5202	Ecil	4	767	3165	3311 4139		TaqI	13						
BbsI	4	1136	1475	1849	2209		Eco47III	3	395	1896	2345		TaqII	9	898	1116	1789	2993 4332		
BbvI	26						Eco57I	2	3639	4651					4517	4670	4687	5031		
BccI	13						EcoNI	1	525				TfiI	5	1669	1971	2141	2645 3066		
Bce83I	7	21	1804	1974	3182	3480	EcoO109I	3	53	423	2097		ThaI	34						
		3721	4589				EcoRI	1	192				TseI	26						
BceII	5	509	850	1477	3593	5153	EcoRII	7	713	1028	1568 1625 3117		Tsp45I	8	1171	1999	2530	2743 2838		
Bcgl	10	160	194	1282	1316	1816			3238	3251					4240	4451	5300			
		1850	2668	2702	4489	4523	EcoRV	1	1440				Tsp509I	15						
BclI	1	1004					FauI	17					Tth111I	1	2836					
Bfal	7	70	203	2105	3586	3839	FokI	10	1036	1045	2310 2372 2450		Tth111II	6	829	1522	2552	3681 3688		
		4174	5278						2636	2777	3950 4131 4418				3720					
BglI	2	2054	4104				FspI	2	2072	4206			UbaJI	20						
BglII	1	268					GdiII	5	166	298	430 1664 4372		VspI	4	251	1675	1734	4156		
BmgI	1	1199					HaeI	5	718	2039	3106 3117 3569		XcmI	3	846	1362	1380			
BpmI	5	828	1317	1951	2618	4054	HaeII	14					XhoI	1	158					
Bpu10I	1	2197					HaeIII	24					XmnI	2	2649	4583				
Bpu1102I	1	80					Hgal	12					Enzymes that do not cut pET-21(+):							
Bsal	1	4045					HgiEII	2	588	3677			AatII	AflII	AgeI	AscI	AvrII			
BsaAI	2	2843	5127				HhaI	45					BaeI	BseRI	BsmI	BspMI	BsrGI			
BsaBI	3	267	273	2288			Hin4I	3	889	3983	4057		Bsu36I	Clal	FseI	KpnI	MscI			
BsaHI	6	313	334	448	947	1630	HincII	2	181	1496			MunI	NcoI	NdeI	NheI	NruI			
		4521					HindIII	1	173				Nsil	NspV	PacI	PmeI	PmlI			
BsaJI	6	57	427	433	1625	2063	Hinfl	14					RleAI	RsrII	SacII	SexAI	SfiI			
		3251					HpaI	1	1496				Sgfl	Smal	SnaBI	SpeI	SrfI			
BsaWI	7	2	1309	1812	2280	3297	HphI	16					Sse8387I	StuI	SunI	Swal	XbaI			
		3444	4275				Maell	15												
BsaXI	2	1649	5075				MaellI	18												
Bsbl	2	2807	5034				MbolI	14												
BscGI	13						MluI	1	990											
BsgI	3	841	1041	2251			Mmel	3	3306	3490	5104									
Bsil	2	3264	4648				MnII	26												
BsiEI	6	169	1775	3007	3431	4354	MseI	26												
		4503					MslI	9	1042	1330	1360 2078 2273									
BsII	20								2664	4236	4395 4754									
BsmAI	7	687	1092	1218	1605	2732	MspI	31												
		4045	4821				MspA1I	9	84	1020	1590 1683 2682									
BsmBI	2	1605	2732						2801	3433	3678 4619									
BsmFI	4	451	1992	2362	5342		Mwol	38												
BsoFI	46						NarI	4	313	334	448 1630									
Bsp24I	10	280	312	831	863	1133	NciI	12												
		1165	3584	3616	3762	3794	NgoAIV	4	300	1888	2048 5228									
Bsp1286I	13						NlaIII	24												
BspEI	2	2	2280				NlaIV	25												
BspGI	1	2617					NotI	1	166											
BspLU11I	1	3091					NspI	4	465	2436	2728 3095									
BsrI	25						Pfl1108I	2	1877	4002										
BsrBI	4	223	3024	4825	5271		PfIMI	1	572											
BsrDI	4	1037	1403	4045	4219		PleI	9	251	539	626 1422 2985									
BsrFI	7	300	309	676	1888	2048			3470	3973	5062 5070									
		4064	5228				PshAI	1	1835											
BssHII	1	1401					Psp5II	1	2097											