禾谷孢囊线虫（*Heterodera avenae*）*Ha34609*基因功能预测

Prediction of Ha34609 Gene Function in *Heterodera avenae*

汇报人：坚晋卓

G08小组成员： 坚晋卓　彭新红

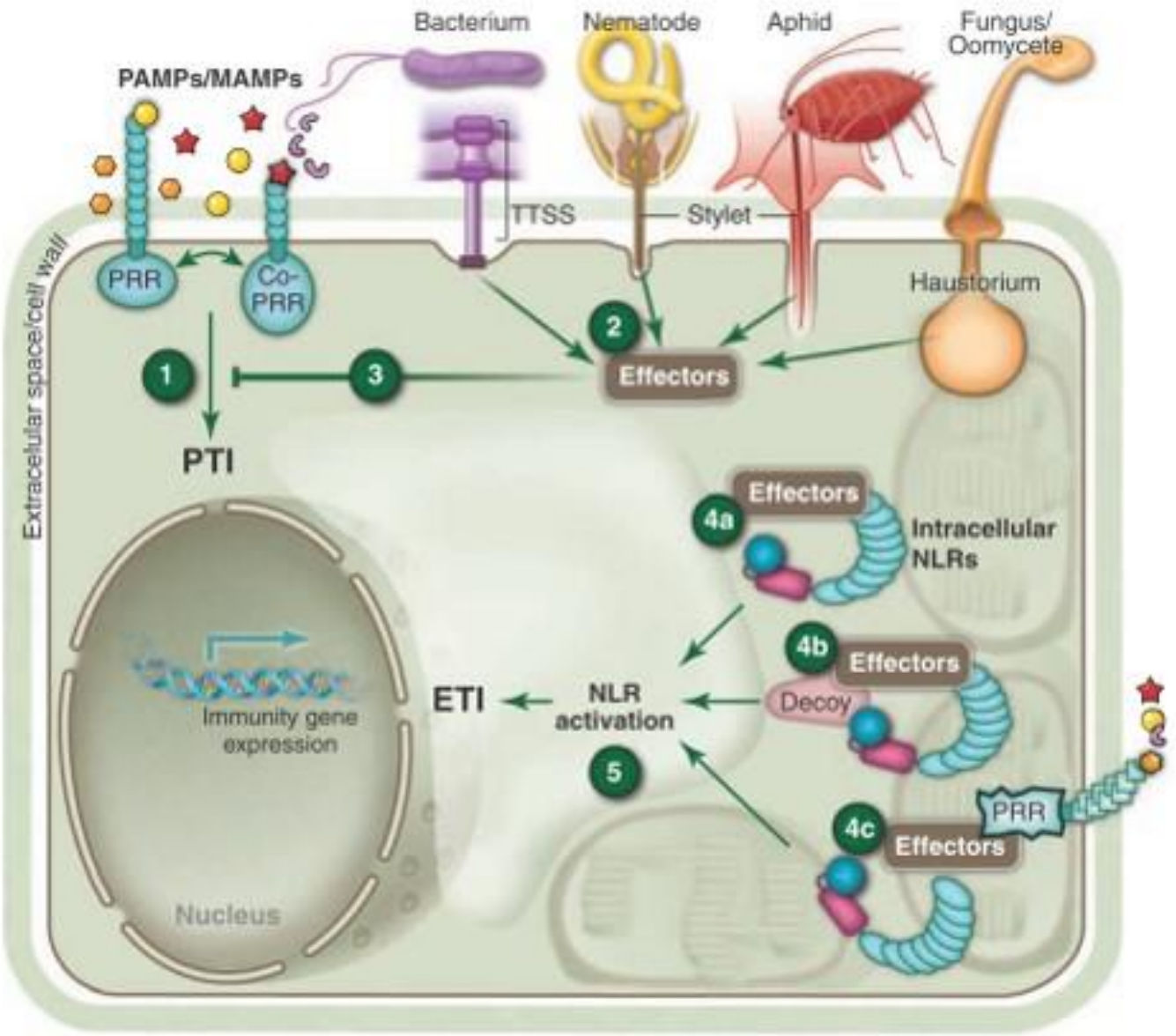
李云卿　闫大琦

汇报时间：2017-12-13

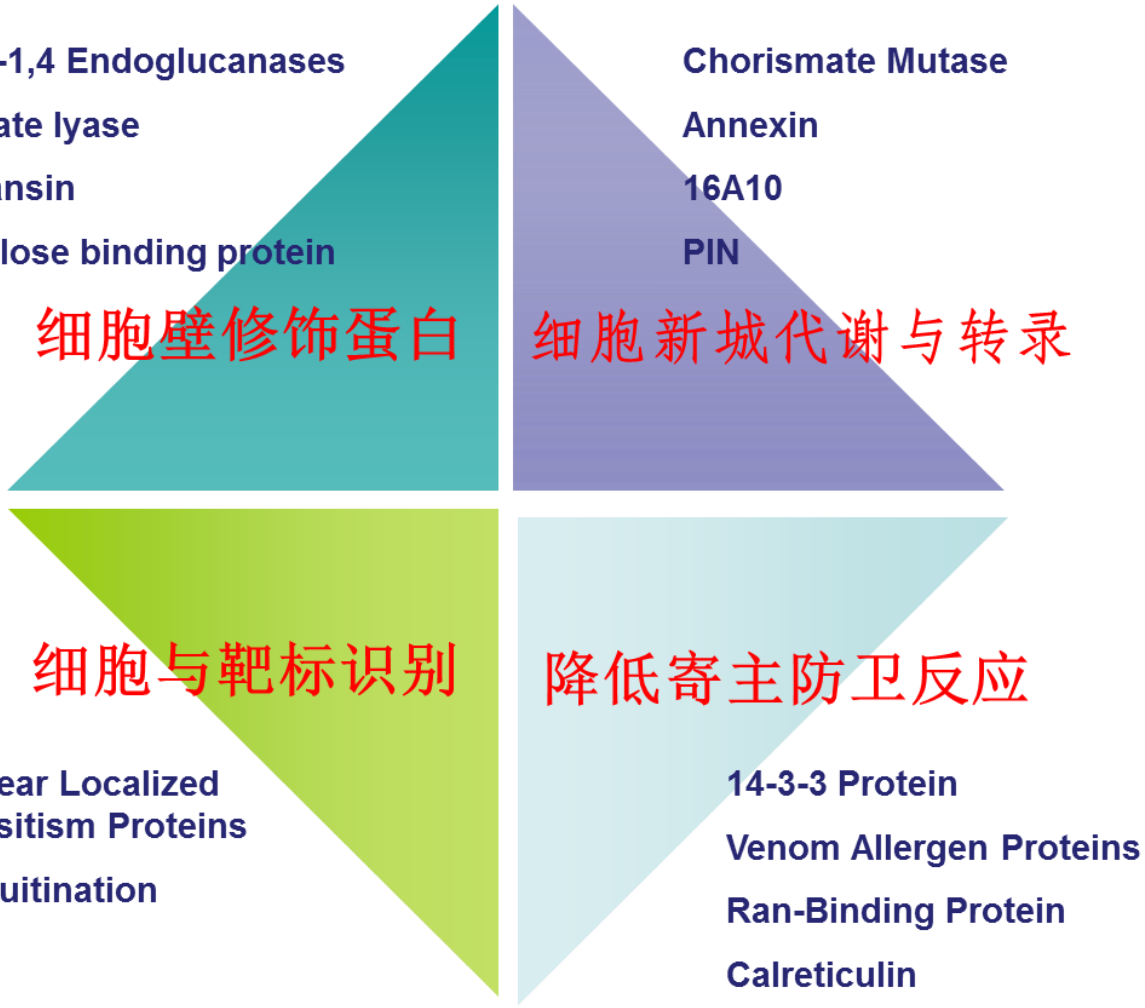
纲　要

1. 研究背景
2. 组员课题相关
3. ORF片段预测
4. Signal peptide预测
5. 跨膜蛋白预测
6. 限制性内切酶位点分析
7. 植物表达载体质粒图谱分析
8. 用Primer5.0引物设计
9. 用DNAMAN检测引物特异性
10. 亚细胞定位的预测

1.研究背景


Bacterium
Nematode
Aphid
Fungus/
Oomycete
PAMPs/MAMPs
TTSS
Stylet
PRR
Co-
PRR
Haustorium
2
Effectors
1
3
PTI
Extracellular space/cell wall
Effectors
4a
Intracellular
NLRs
4b
Effectors
Decoy
Immunity gene
expression
ETI
NLR
activation
5
PRR
4c
Effectors
Nucleus

1.研究背景


Beta-1,4 Endoglucanases
pectate lyase
Expansin
cellulose binding protein
细胞壁修饰蛋白
Chorismate Mutase
Annexin
16A10
PIN
细胞新城代谢与转录
细胞与靶标识别
Nuclear Localized
Parasitism Proteins
Ubiquitination
降低寄主防卫反应
14-3-3 Protein
Venom Allergen Proteins
Ran-Binding Protein
Calreticulin

2.组员课题相关

G08D_李云卿

棉花抗黄萎病相关基因的生物信息学分析及功能验证

对棉花抗病品种接种黄萎病菌VD080，在不同时间段取样、分析，筛选差异表达基因，并对候选基因进行后续的生物信息学分析和功能验证。

以候选基因Gh_A06G1627为例进行结构和功能分析。

Content

- 找相似序列——Blast
- 多序列比对——MEGA7
- 建进化树——MEGA7
- CDs序列转换成氨基酸序列——EMBOSS Transeq
- 分析蛋白结构——Protscale
- 预测蛋白模型——Phy2（ Swiss –pdb Viewer ）
- 预测功能模块——SMART
- 预测蛋白——Predict Protein
- 预测互做蛋白——STRING
- 引物设计——Primer 5

2.组员课题相关

G08A_彭新红

哈茨木霉Th-33 thga3 基因的功能研究

研究目的及意义：

G蛋白信号系统在调控真菌的生长、产孢、拮抗和次生代谢等生物学过程中发挥着重要作用。GaIII亚基能调控真菌细胞壁相关酶活性，从而影响真菌的重寄生和拮抗能力，同时它还能影响真菌的产孢。因此，我们的研究拟采用基因敲除、回复突变技术，结合转录组测序数据分析，研究哈茨木霉GaIII亚基*thga3* 基因对木霉菌株生长、产孢、重寄生作用的影响，从而明确该基因的功能。

2.组员课题相关

可能用到的相关软件

- 利用Primer5软件设计引物
- 利用DNAMAN软件和NCBI中ORF finder分析*thga3*的开放阅读框和Blast比对
- 利用ExPASy中的protScale分析编码蛋白质的疏水性
- 利用ExPASy中的Protparam进行氨基酸理化性质分析
- 利用signal 4.1分析翻译后信号肽
- 利用ExPASy中的TMPRED和 EMBOSS Explorer中的Tmap分析跨膜结构
- 利用MEGA 7.0构建系统发育树
- EMBOSS Explorer中的Rmap分析酶切位点。
- 利用Phyre2 和Swiss-PdbViewer预测和分析蛋白模型

2.组员课题相关

G08A_闫大琦

WRKYs基因功能研究

研究背景

植物具有复杂的基因表达调控体系来抵御病原物的侵染，主要体现为转录因子在转录水平上对各种信号传导途径的调控。转录因子WRKY是一类与植物抗病防卫反应相关的转录因子，为植物中特有的超基因家族，该家族成员都含有高度保守的 WRKYGQK结构域和C2HH/C锌指结构基序，并通过与顺式作用元件W-box:(T)(T)TGAC(C/T)结合来调控靶基因的表达。已有大量研究证实，WRKY基因在植物的防卫反应中起着重要的调控作用。

2.组员课题相关

- 利用Blast找相似序列
- 利用DNAMAN软件比对序列
- 利用ExPASy中的protScale分析编码蛋白质的疏水性
- 利用signal 4.1分析翻译后信号肽
- EMBOSS Explorer中的Tmap分析跨膜结构
- 利用MEGA 7.0构建系统发育树
- 利用Phyre2 预测和分析蛋白模型

G08B_坚晋卓

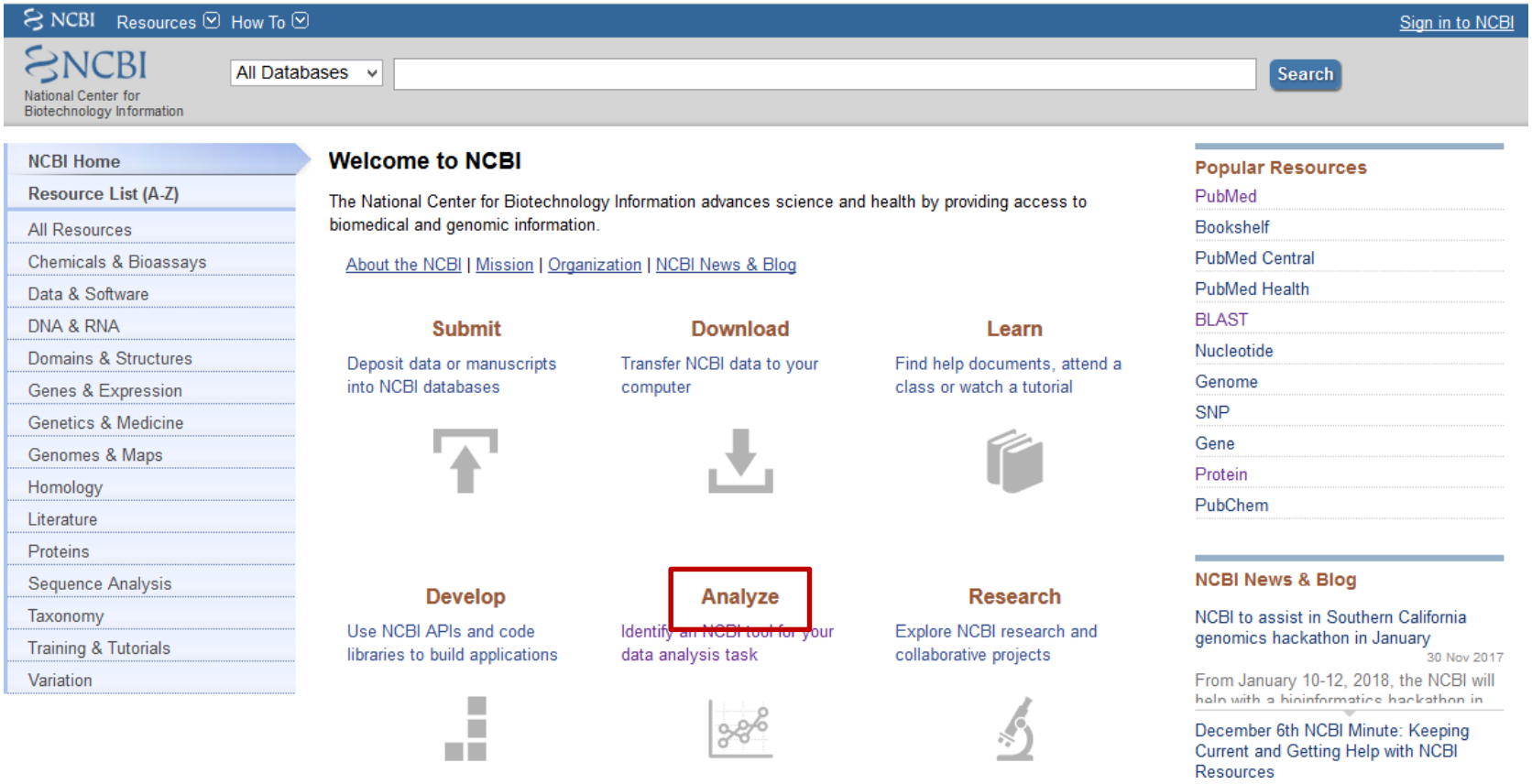
禾谷孢囊线虫（*Heterodera avenae*）*Ha34609*基因功能研究

植物寄生线虫效应子由线虫腺体细胞分泌产生，通过口针将这些蛋白质分泌到寄主细胞中从而促进自身的寄生作用。这些基因主要是在线虫分泌腺细胞中特异性表达，且多数具有信号肽，不含跨膜结构域。

本实验从实验室已构建的禾谷孢囊线虫转录组数据中，克隆出推定效应因子*Ha34609*，效应子对于线虫寄生作用至关重要，但其功能和致病机理我们现在却知之甚少。采用亚细胞定位研究该基因在植物体内的表达部位，以此完成禾谷孢囊线虫*Ha34609*基因的表达特性分析。

3.ORF片段预测

开放性阅读框分析：利用NCBI的ORF Finder程序对cDNA序列做开放性阅读框分析



Analyze

NCBI provides a wide variety of data analysis tools that allow users to manipulate, align, visualize and evaluate biological data.

Selected Analysis Tools

(CDART)	
Conserved Domain Search Service (CD Search)	Identifies the conserved domains present in a protein sequence
Digital Differential Display (DDD)	Identifies genes with significantly different expression levels by comparing EST profiles
Electronic PCR (e-PCR)	Identifies sequence tagged sites (STSs) within DNA sequences
Frequency-weighted Link (FLink)	Links a group of records in a source database to a ranked list of associated records in a destination database based on frequency-weighted statistics
Gene Expression Omnibus (GEO) BLAST	Finds regions of local similarity between query sequences and GenBank sequences included on microarray or SAGE platforms in the GEO database
Genetic Codes	Displays the genetic codes for organisms in the Taxonomy database in tables and on a taxonomic tree
Genome BLAST	Finds regions of local similarity between query sequences and genome sequences
Genome ProtMap	Maps each protein from a COG or VOG back to its genome
Genome Remapping Service	Projects genome annotation data from one genomic assembly to another
Genome Workbench	Integrated application for viewing and analyzing sequence data
LinkOut	Allows third parties to link directly from PubMed abstracts and other Entrez database records to relevant web-accessible resources beyond the Entrez system
Map Viewer	Provides special browsing capabilities for genomic maps and assembled sequences for a subset of organisms
Open Reading Frame Finder (ORF Finder)	Suggests potential open reading frames in a DNA sequence
OSIRIS	Facilitates the assessment of multiplex short tandem repeat (STR) DNA profiles based on laboratory-specific protocols
Phenotype-Genotype Integrator (PheGenI)	Finds human phenotype/genotype relationships with queries by phenotype, chromosome location, gene, and SNP identifiers.
Primer-BLAST	Uses Primer3 to design PCR primers to a sequence template
ProSplign	Computes alignments of proteins to genomic nucleotide sequences
PSSM Viewer	Displays and manipulates PSSM matrices from CDD records and PSI-BLAST
PubChem Standardization Service	Processes chemical structures into a standard representation used by PubChem
PubChem Structure Search	Searches PubChem Compound by chemical structure, SMILES, InChI or Molecular Formula

Open Reading Frame Finder

ORF finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF finder to search newly sequenced DNA for potential protein encoding segments, verify predicted protein using newly developed SMART BLAST or regular BLASTP.

This web version of the ORF finder is limited to the subrange of the query sequence up to 50 kb long. Stand-alone version, which doesn't have query sequence length limitation, is available for Linux x64.

Examples (click to set values, then click Submit button) :

- NC_011604 Salmonella enterica plasmid pWES-1; genetic code: 11; 'ATG' and alternative initiation codons; minimal ORF length: 300 nt
- NM_000059; genetic code: 1; start codon: 'ATG only'; minimal ORF length: 150 nt

Enter Query Sequence

Enter accession number, gi, or nucleotide sequence in FASTA format:

```
gaaaacaacacatttgttcattaattaggagtttaaattcaatttaaggcaattaatttgctttaaattaaaa
ATGAGTTTATCAATATTTTTGTGGTTGTTCACCATCGTTTCCTTTACTACATCCGCCGCTTCACATAAATCGTATCAACAAC
aataattggaaatgtgaatgtcttataattacttttttccttttttgtgctttaatttttttggaaatattcgttttatgatgaa
```

Choose Search Parameters

Minimal ORF length (nt): 75

Genetic code: 1. Standard

ORF start codon to use:

- (•) "ATG" only
- ( ) "ATG" and alternative initiation codons
- ( ) Any sense codon

Ignore nested ORFs: ☐

Start Search / Clear

Submit Clear

3.ORF片段预测

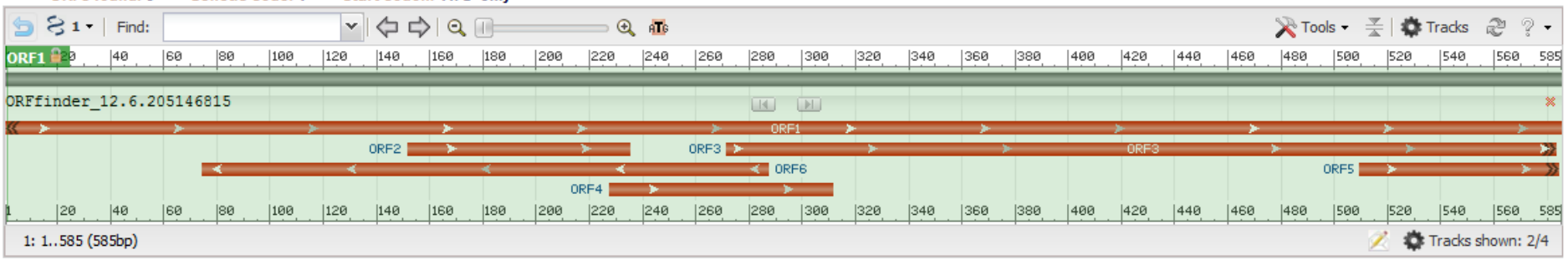
NCBI Resources How To Sign in to NCBI

ORFfinder PubMed Search

Open Reading Frame Viewer

Sequence

ORFs found: 6 Genetic code: 1 Start codon: 'ATG' only



Six-frame translation...

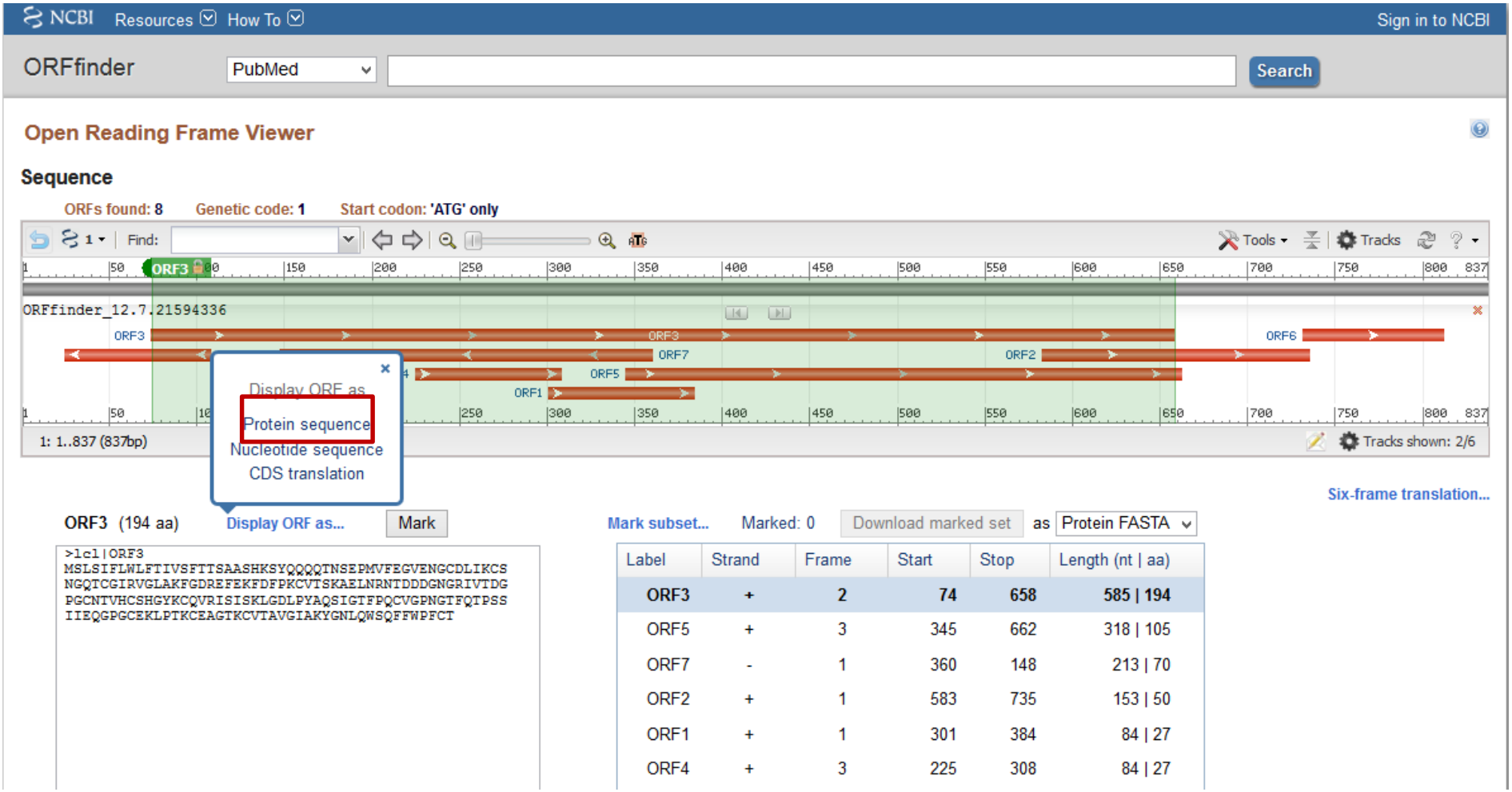
ORF1 (585 nt) Display ORF as... Mark

Mark subset... Marked: 0 Download marked set as Protein FASTA

```
>lcl|ORF1 CDS
ATGAGTTTATCAATATTTTTGTGGTTGTTCACCATCGTTTCCTTTACTAC
ATCCGCCGCTTCACATAAATCGTATCAACAACAGCAAACAAACAGTGAGC
CGATGGTATTTGAAGGCGTTGAAAATGGGTGCGATTTAATCAAATGTTCA
AATGGGCAAACTTGCGGTATTCGAGTTGGTCTTGCAAAATTTGGCGACAG
AGAATTTGAGAAATTTGATTTTCCAAAATGTGTGACAAGCAAAGCGGAAT
TGAACCGAAACACGGACGACGATGGAAATGGGCGCATTGTCACGGATGGA
CCAGGATGTAATACTGTACACTGTAGCCACGGTTACAAATGCCAAGTGCG
CATTTCAATTTCTAAGCTTGGCGATTTACCGTATGCCCAATCAATCGGAA
CCTTCCCTCAATGTGTGGGTCCAAATGGCACATTTCAAACGCCAAGTTCA
ATTATCGAACAAGGCCCTGGGTGTGAAAAACTTCCGACTAAATGTGAAGC
AGGCACAAAATGTGTCACCGCTGTTGGAATTGCGAAATATGGAAATTTGC
AATGGTCTCAGTTCTTCTGGCCATTTTGCACTTGA
```

Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF1	**+**	**1**	**<1**	**585**	**585 | 194**
ORF3	+	2	272	>583	312 | 103
ORF6	-	2	287	75	213 | 70
ORF2	+	2	152	235	84 | 27
ORF4	+	3	228	311	84 | 27
ORF5	+	3	510	>584	75 | 24

3.ORF片段预测


NCBI Resources How To
Sign in to NCBI
ORFfinder
PubMed
Search
Open Reading Frame Viewer
Sequence
ORFs found: 8
Genetic code: 1
Start codon: 'ATG' only
Find:
Tools
Tracks
ORFfinder_12.7.21594336
ORF3
ORF7
ORF5
ORF1
ORF6
ORF2
Display ORF as
Protein sequence
Nucleotide sequence
CDS translation
1: 1..837 (837bp)
Tracks shown: 2/6
Six-frame translation...
ORF3 (194 aa)
Display ORF as...
Mark
Mark subset...
Marked: 0
Download marked set
as
Protein FASTA
>lcl|ORF3
MSLSIFLWLFTIVSFTTSAASHKSYQQQQTNSEPMVFEGVENGCDLIKCS
NGQTCGIRVGLAKFGDREFEKFDFPKCVTSKAELNRNTDDDGNGRIVTDG
PGCNTVHCSHGYKCQVRISISKLGDLPYAQSIGTFPQCVGPNGTFQTPSS
IIEQGPGCEKLPTKCEAGTKCVTAVGIAKYGNLQWSQFFWPFCT
Label
Strand
Frame
Start
Stop
Length (nt | aa)
ORF3 + 2 74 658 585 | 194
ORF5 + 3 345 662 318 | 105
ORF7 - 1 360 148 213 | 70
ORF2 + 1 583 735 153 | 50
ORF1 + 1 301 384 84 | 27
ORF4 + 3 225 308 84 | 27

4.Signal peptide（信号肽）预测

是引导新合成的蛋白质向分泌通路转移的短(长度5-30个氨基酸)肽链

打开abc

↓

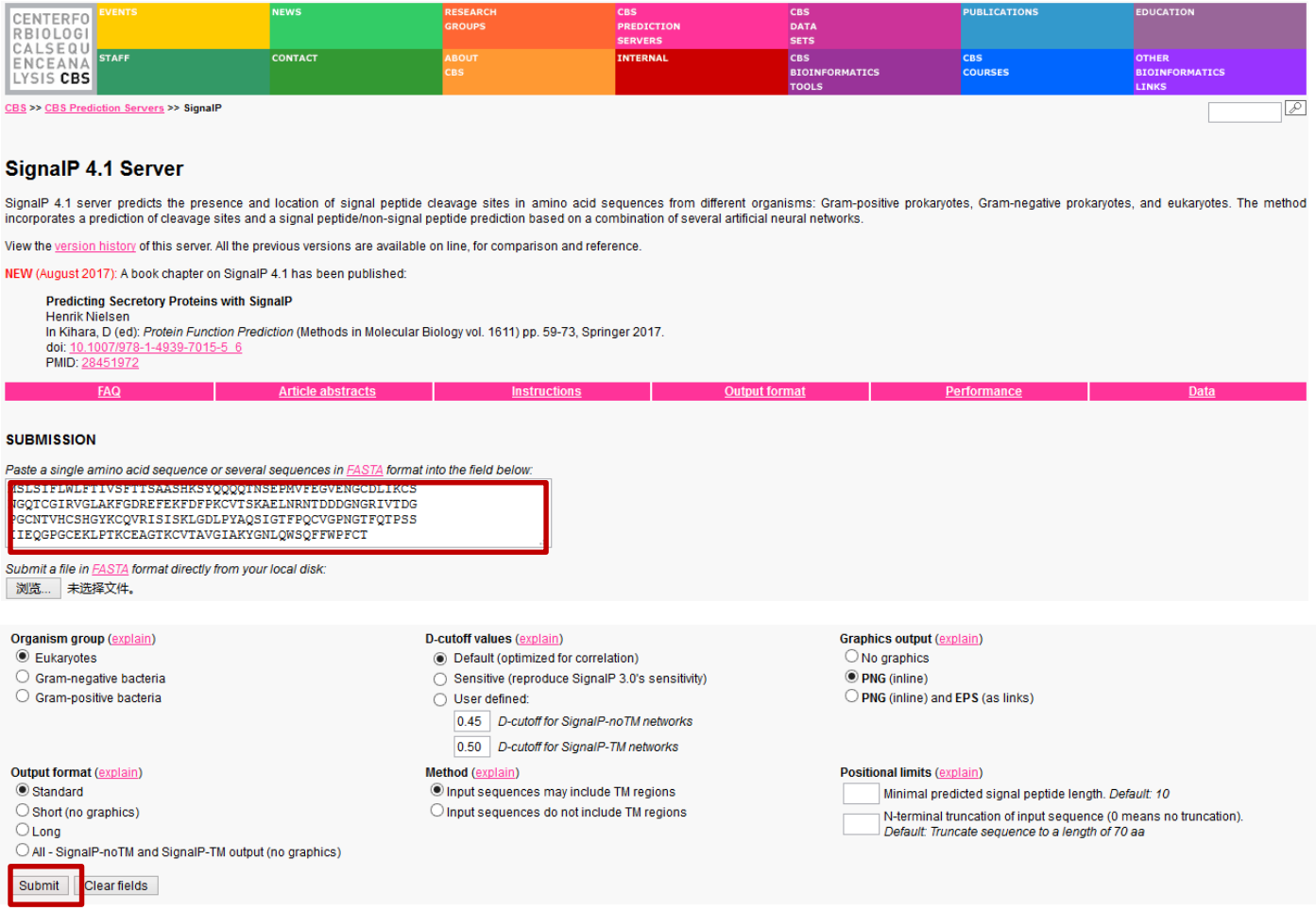
TOOLS

↓

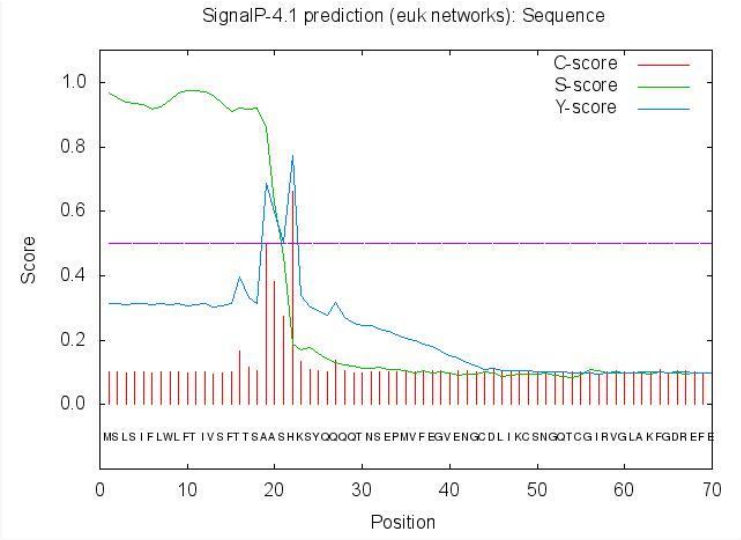
找到[CBS tools]

↓

SignalP 4.1 Server



4.Signal peptide预测




```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    22        0.661
  max. Y    22        0.773
  max. S    11        0.976
  mean S     1-21     0.900
       D     1-21     0.841   0.450   YES
Name=Sequence   SP='YES' Cleavage site between pos. 21 and 22: AAS-HK D=0.841 D-cutoff=0.450 Networks=Signal
# data
# gnuplot script
```

```
Signal sequence prediction with scores over 3.50


>lcl|ORF3
MSLSIFLWLFTIVSFTTSAASHKSYQQQQTNSEPMVFEGVENGCDLIKCS
NGQ from 1 to 70


Maximum score 6.6 at 19

 Signal Sequence: MSLSIFLWLFTIVSFTTS
 Mature Sequence: AASHKSYQQQQTNSEPMVFEGVENGCDLIK...



Other possible sites

Score 6.0 at 20

Score 5.1 at 21

Score 4.0 at 17

Score 3.7 at 15
```

5.跨膜结构预测

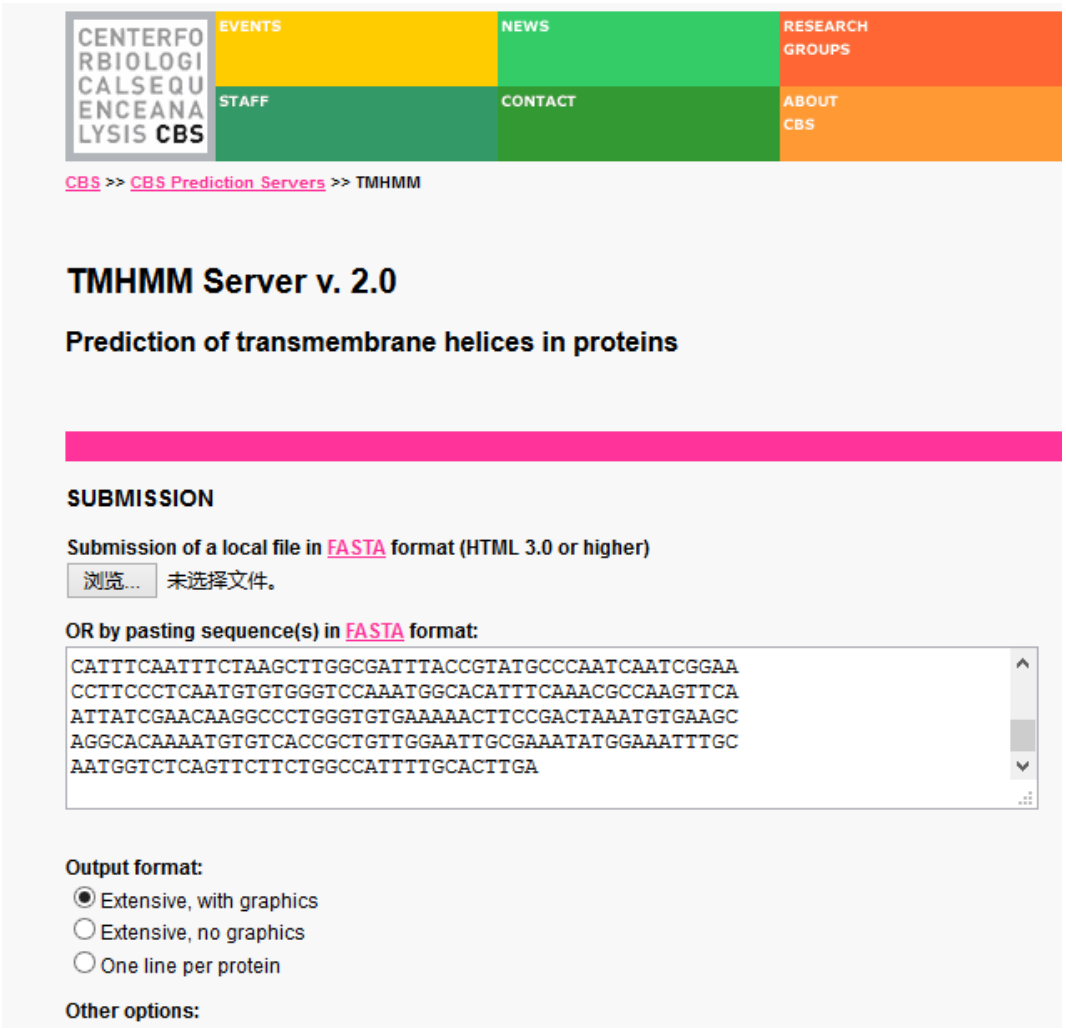
打开abc

⬇

Tools

⬇

TMHMM

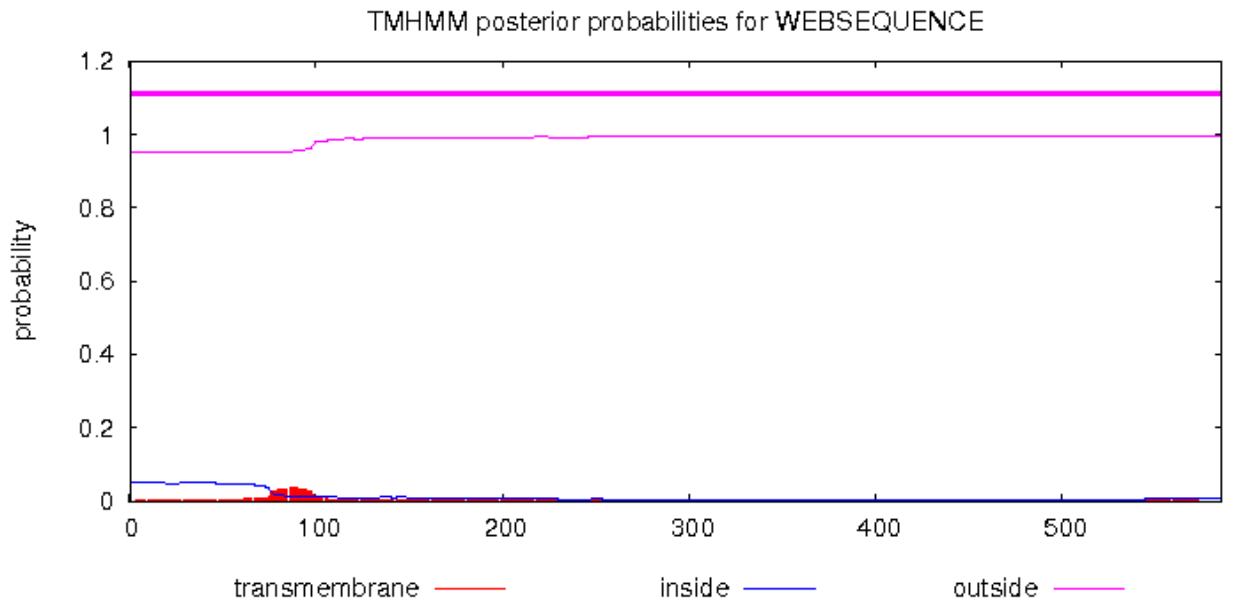


5.跨膜结构预测

TMHMM result

HELP with output formats

```
# WEBSEQUENCE Length: 585
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 1.44465
# WEBSEQUENCE Exp number, first 60 AAs:  0.10711
# WEBSEQUENCE Total prob of N-in:        0.04904
WEBSEQUENCE     TMHMM2.0        outside      1   585
```

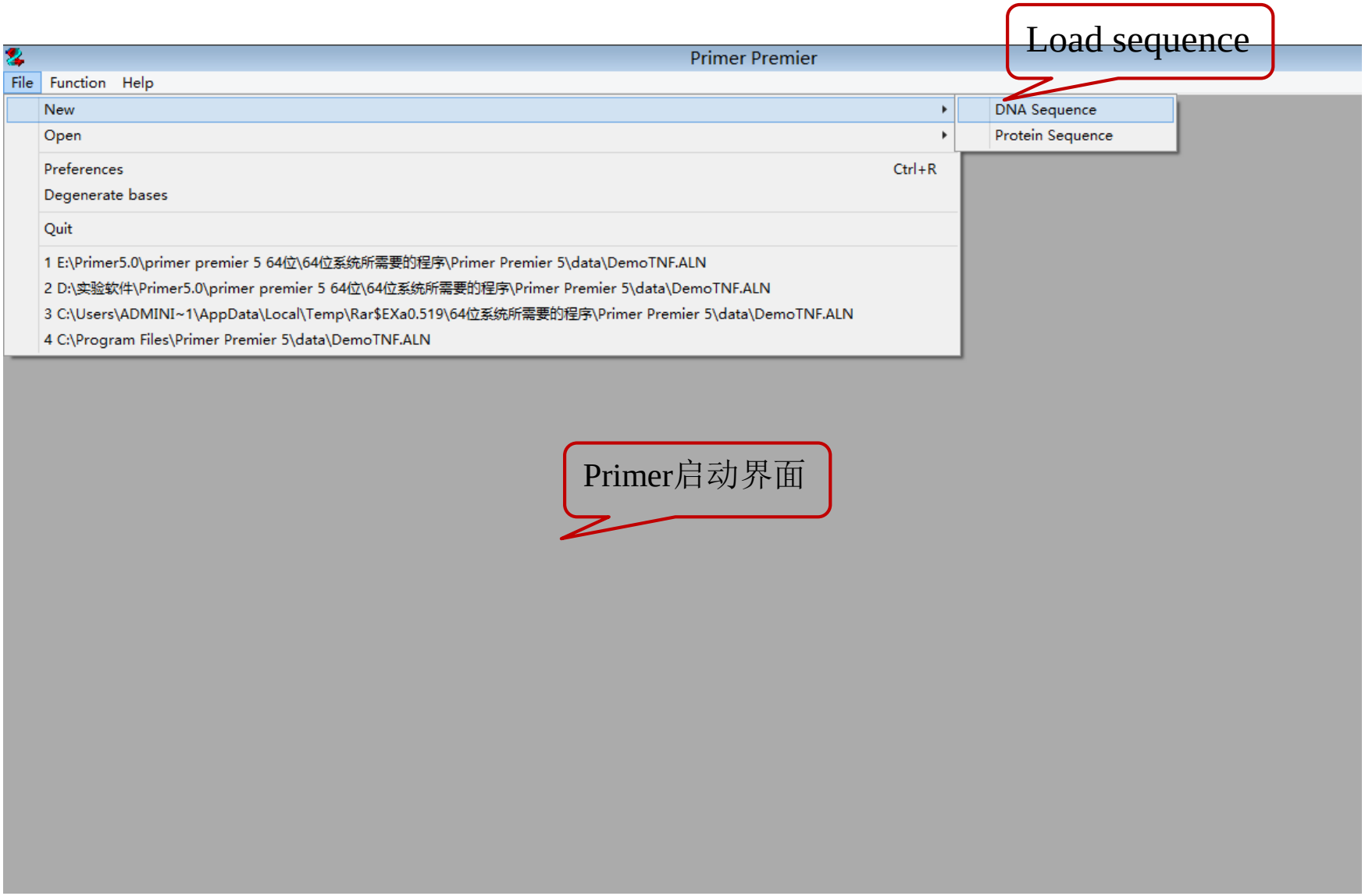


plot in postscript, script for making the plot in gnuplot, data for plot

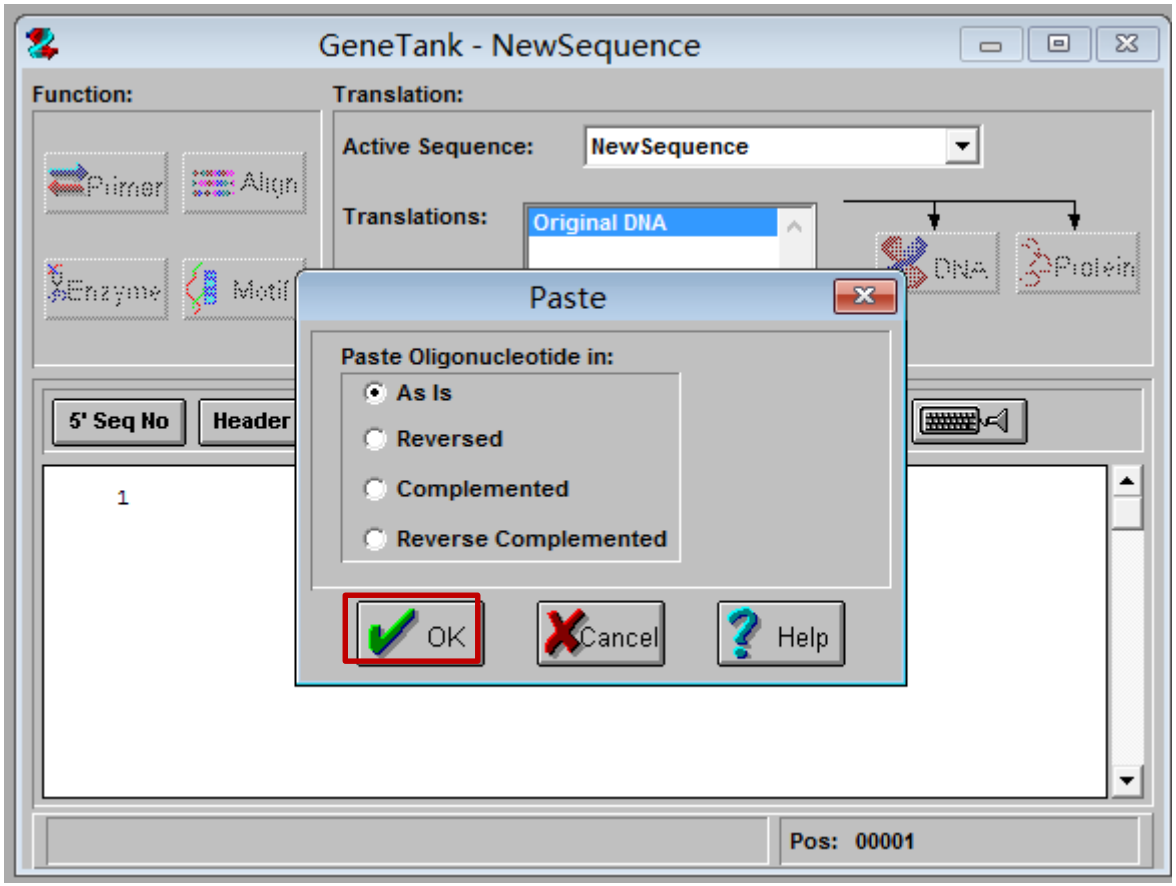
不具有跨膜结构域，说明该基因产物可能被线虫分泌到细胞外部

6.限制性内切酶位点分析

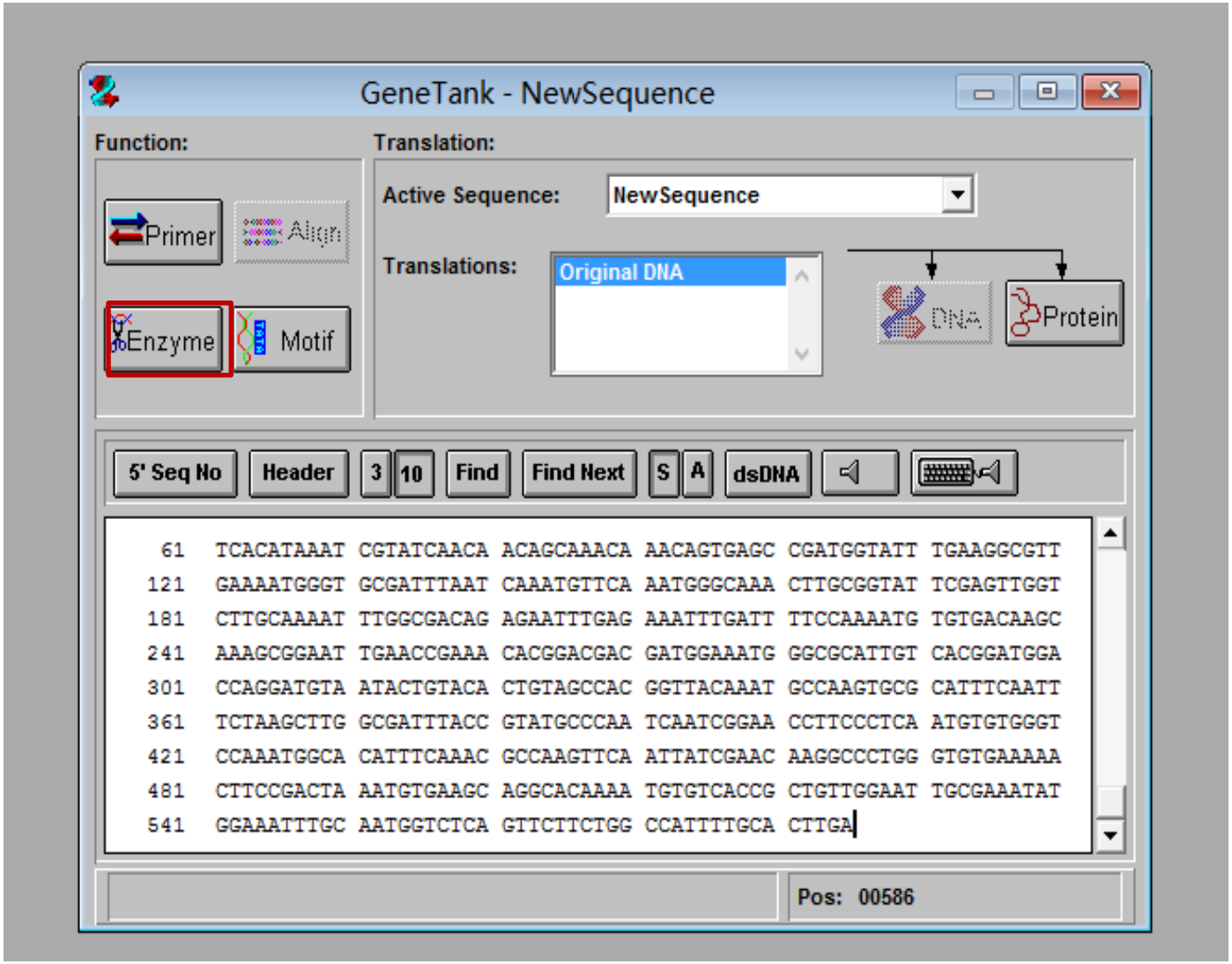
用primer5.0去分析目的片段上的酶切位点



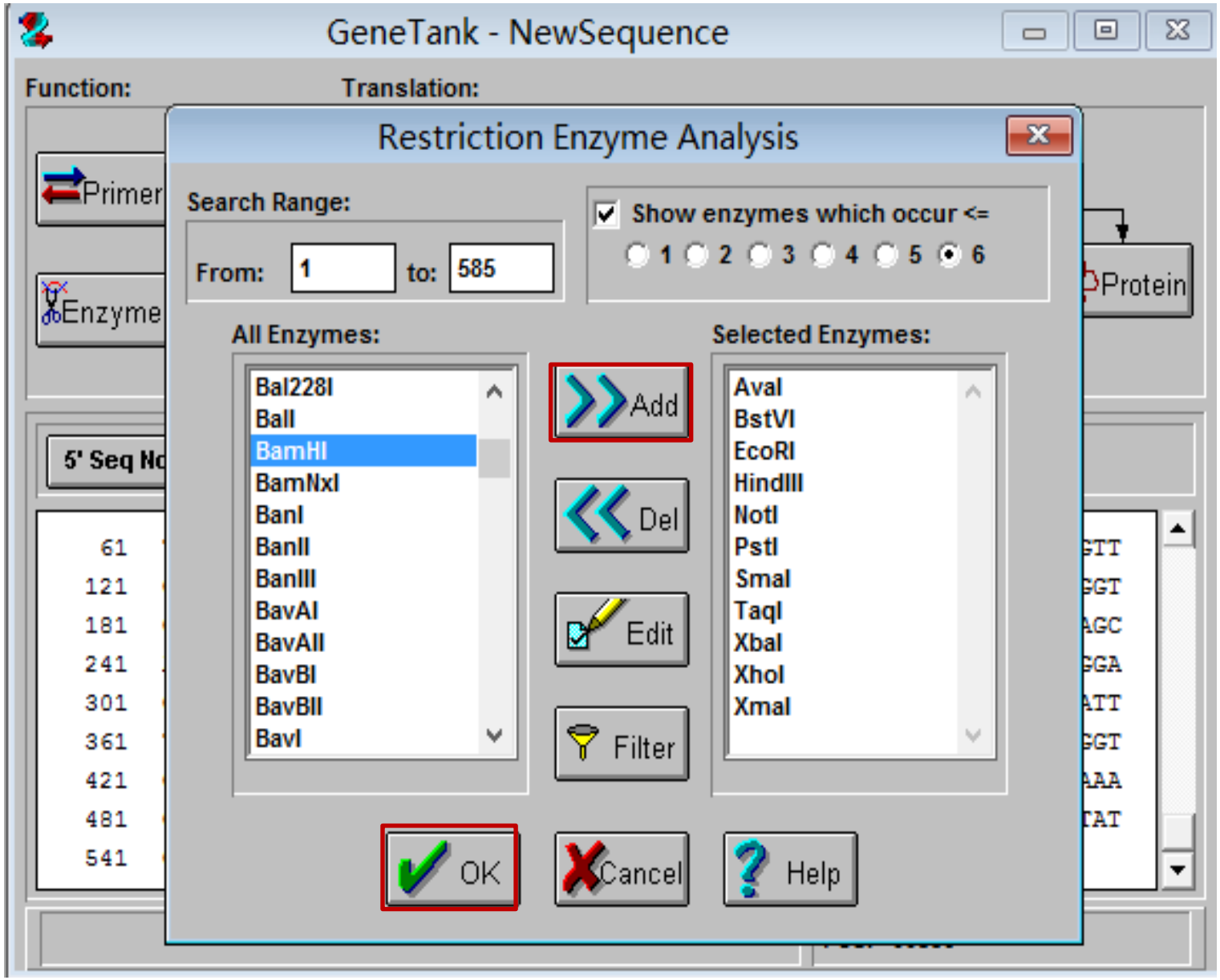
6.限制性内切酶位点分析



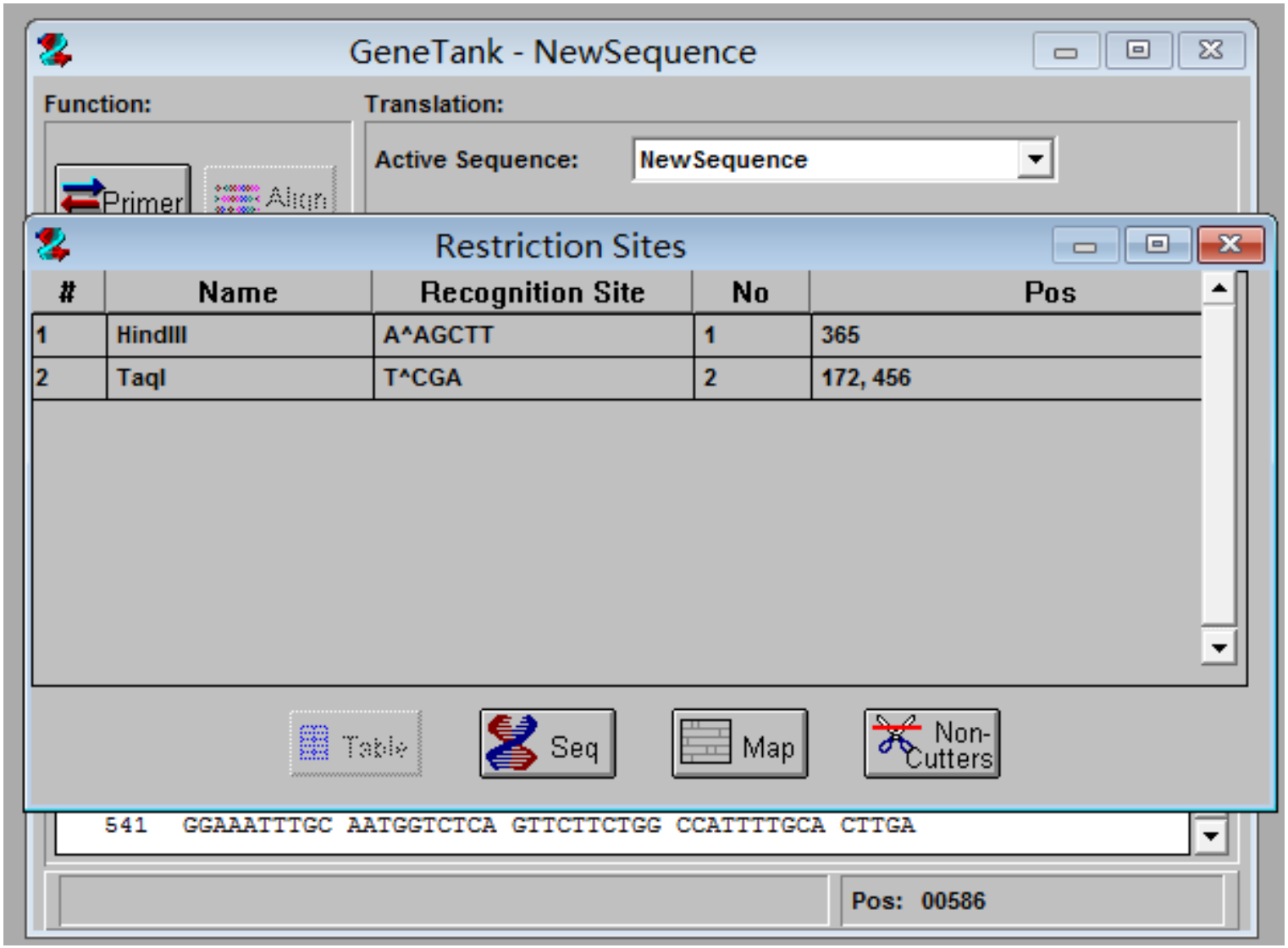
6.限制性内切酶位点分析



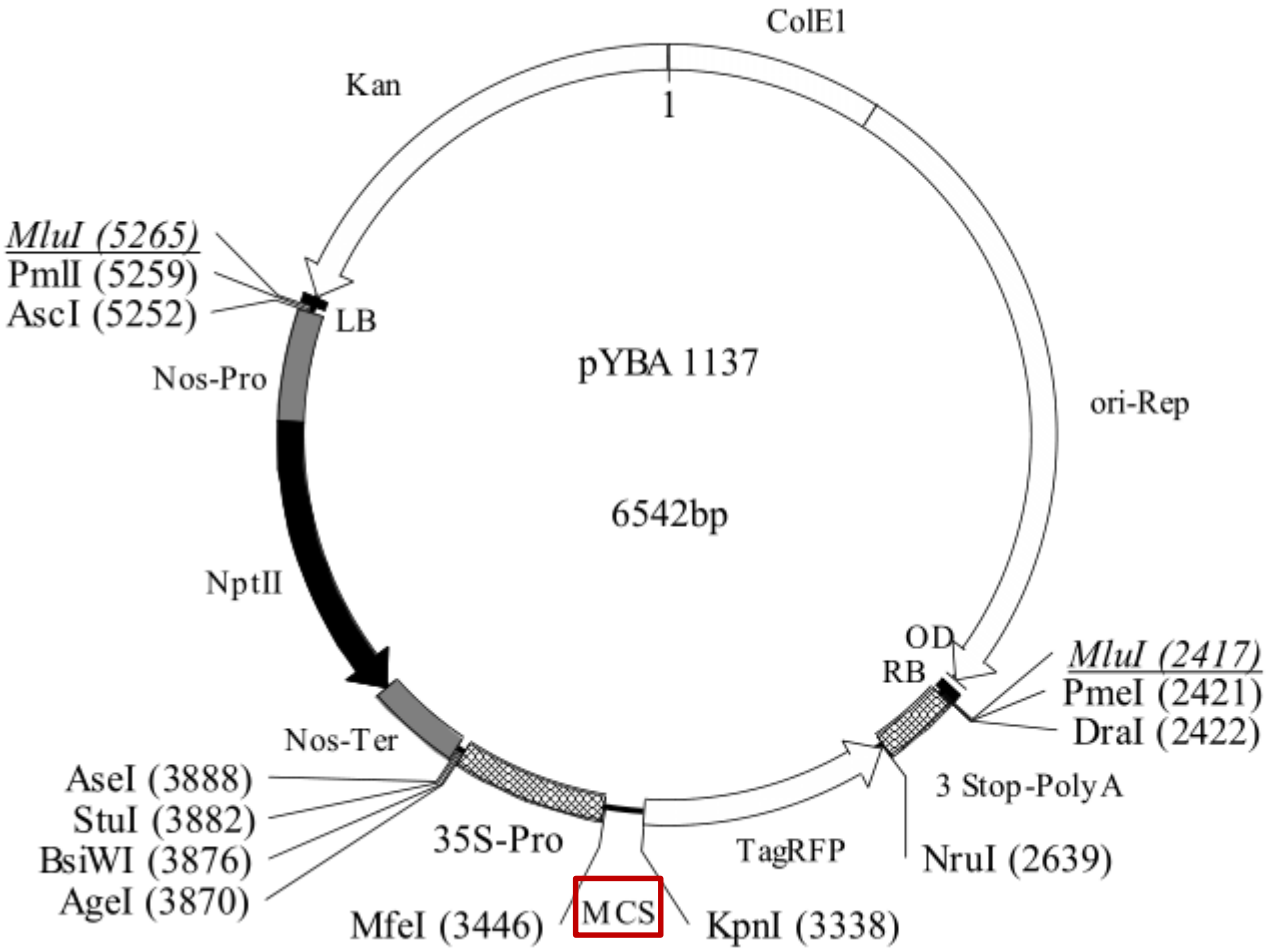
6.限制性内切酶位点分析



6.限制性内切酶位点分析



7. 植物表达载体质粒图谱分析



***Kpn*I（*Acc*65I）**
***Apa*I（*Psp*OMI）**
***Eco*O109I**
***Xho*I**
***Psp*XI**
***Sal*I**
***Cla*I**
***Hin*dIII**
***Eco*RV**
***Eco*RI**
***Pst*I**
***Sma*I（*Xma*I）**
***Bam*HI**
***Spe*I**
***Xba*I**
***Not*I**
***Ale*I**
***Sac*I**
***Mfe*I**

7. 植物表达载体质粒图谱分析

7. 植物表达载体质粒图谱分析

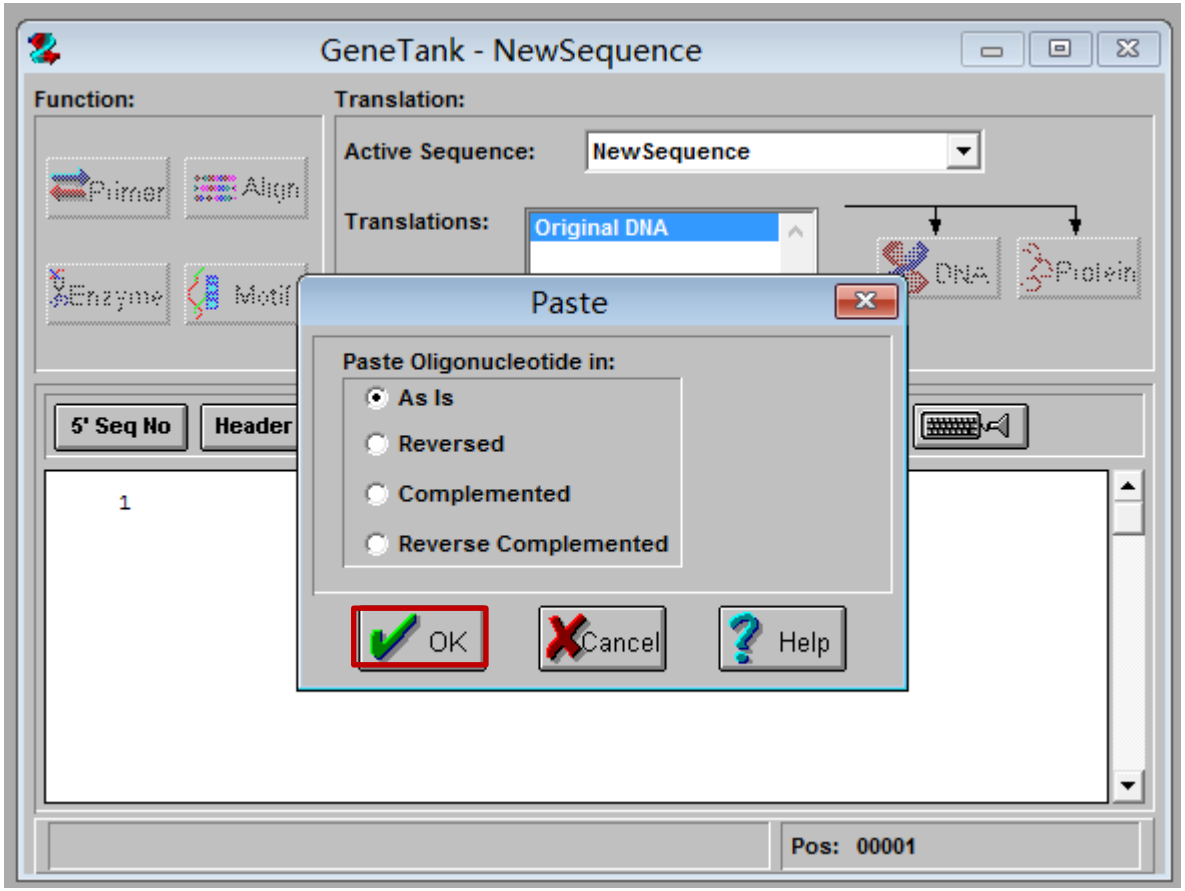
PCR 设计引物时酶切位点的保护碱基表 2

酶	寡核苷酸序列
BamH I	CGGATCCG CGGGATCCCG CGCGGATCCGCG
Xho I	CCTCGAGG CCCTCGAGGG CCGCTCGAGCGG

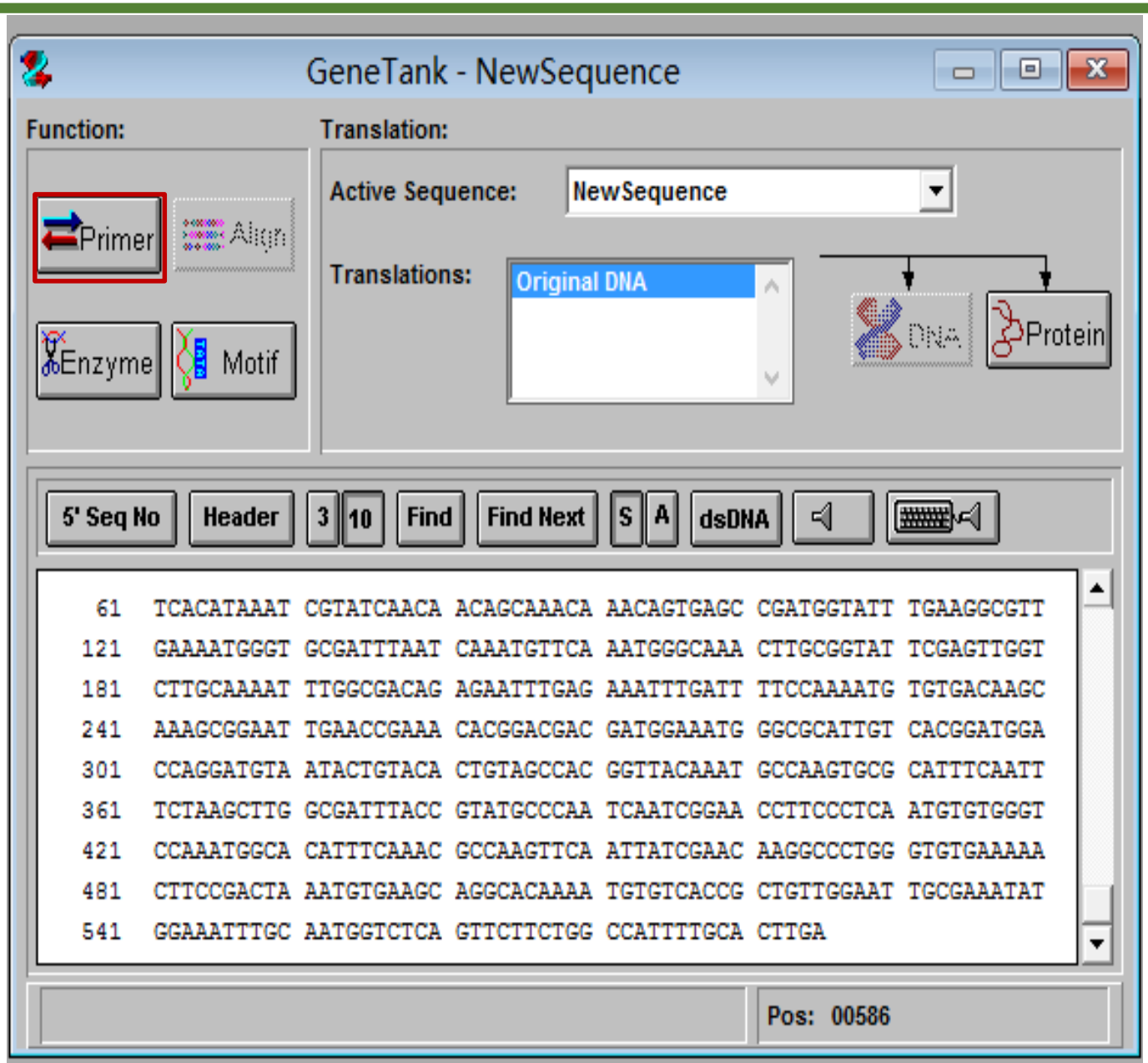
引物的结构就是（5→3）：

保护碱基+酶切位点+引物序列

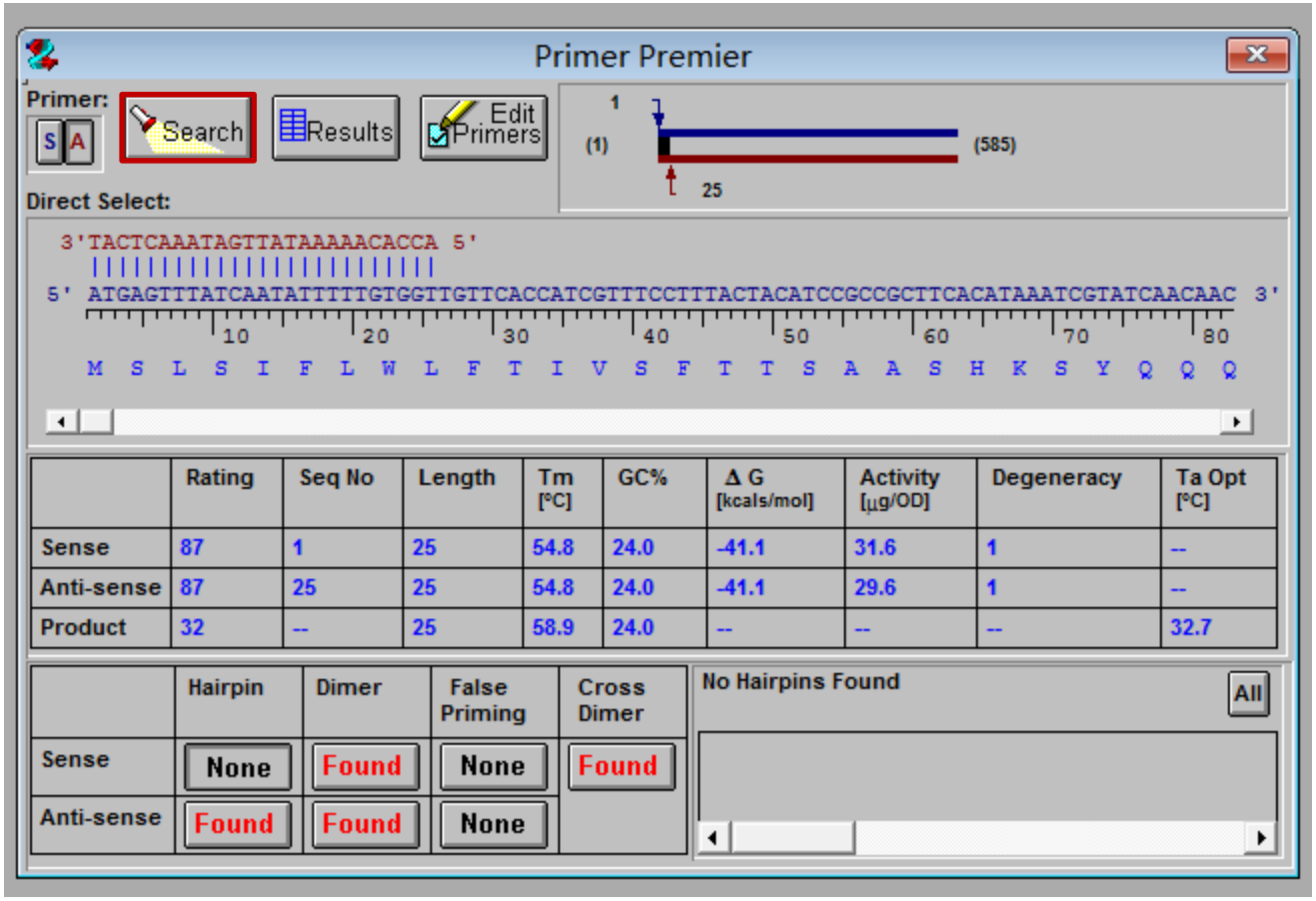
8. 用Primer5.0引物设计



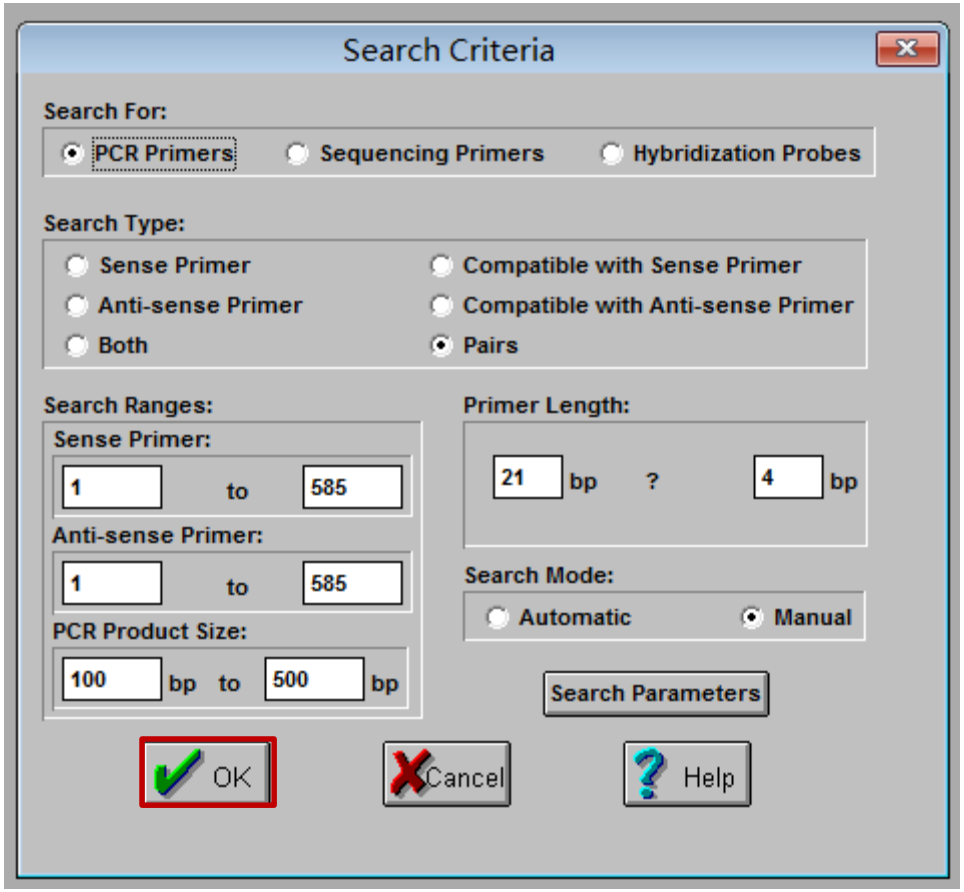
8. 用Primer5.0引物设计



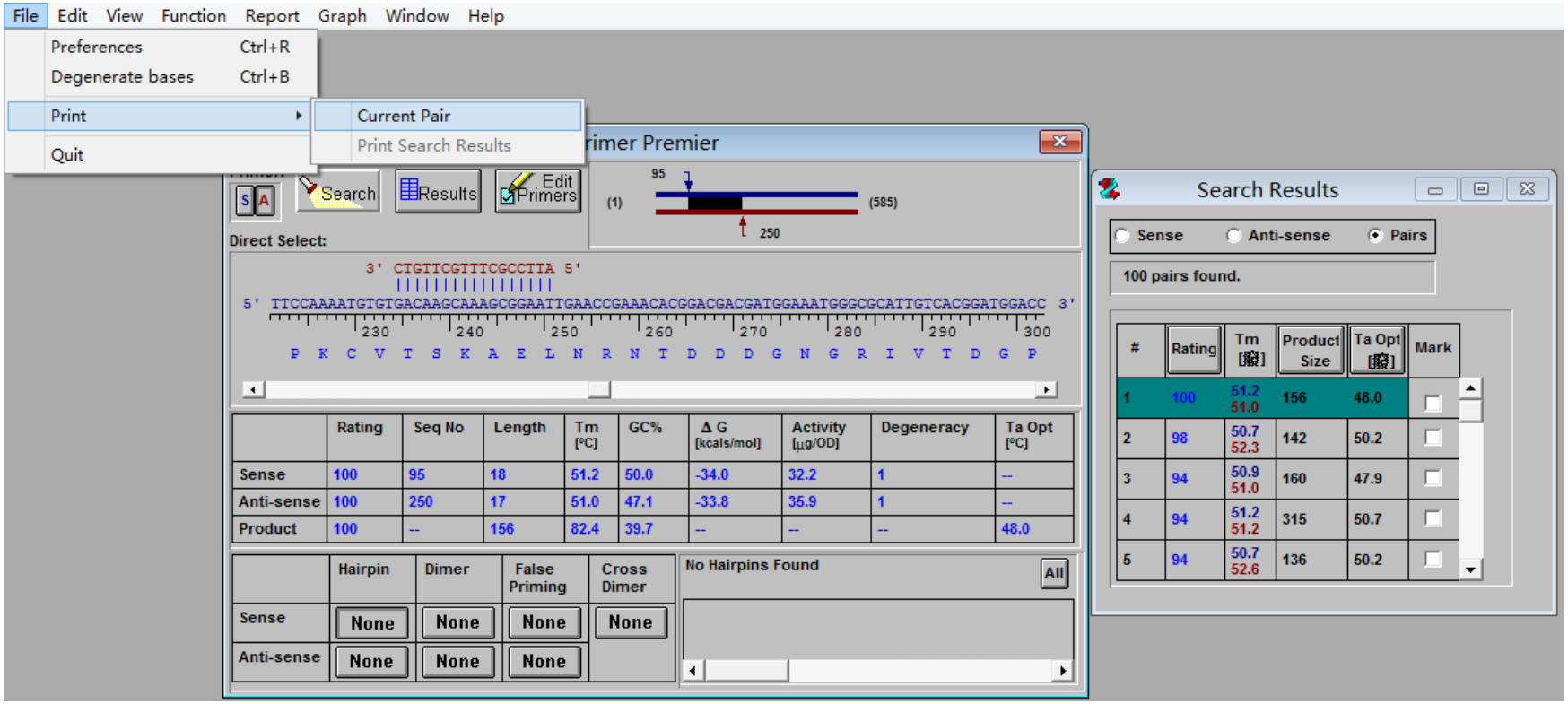
8. 用Primer5.0引物设计



8. 用Primer5.0引物设计

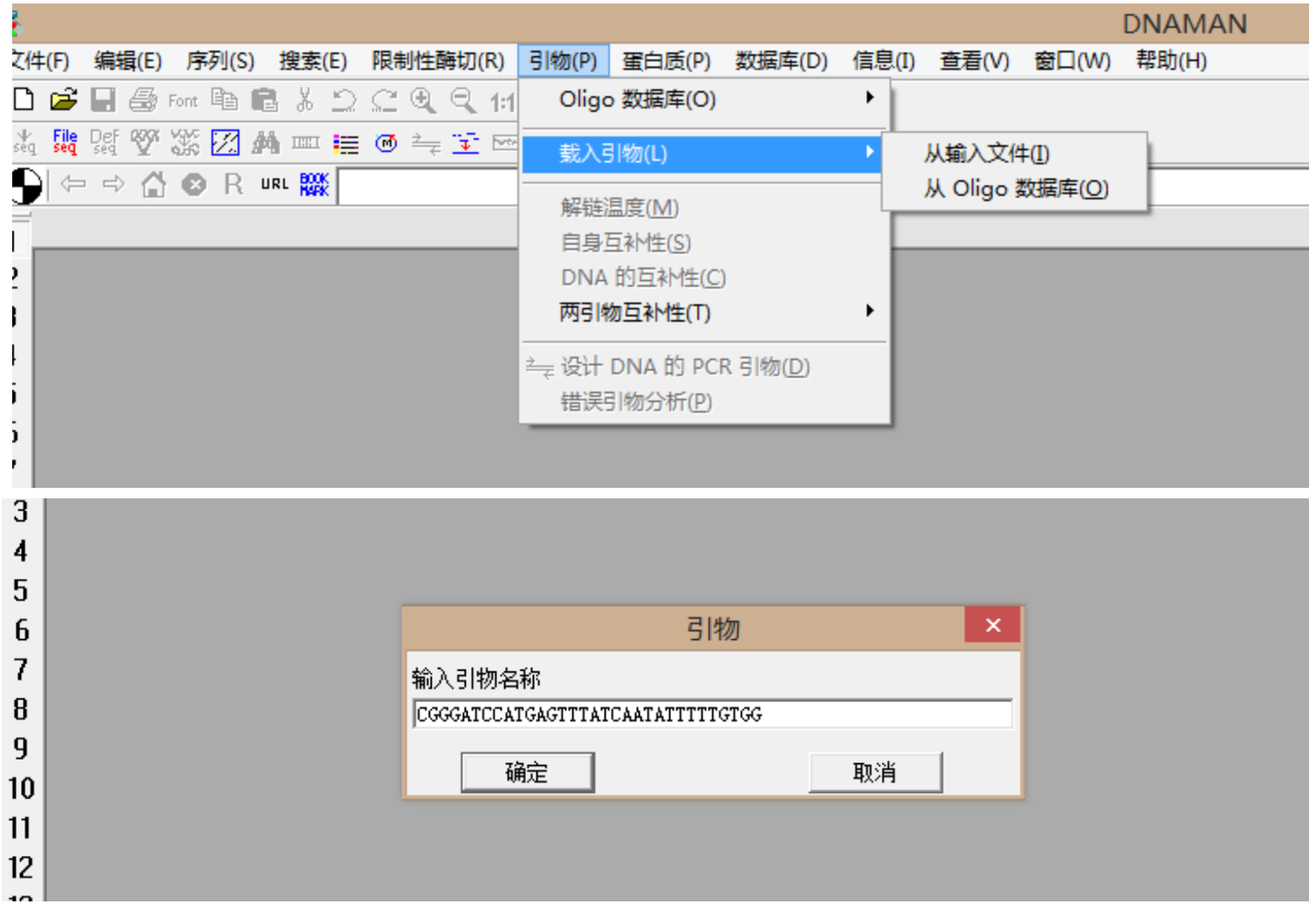


8. 用Primer5.0引物设计

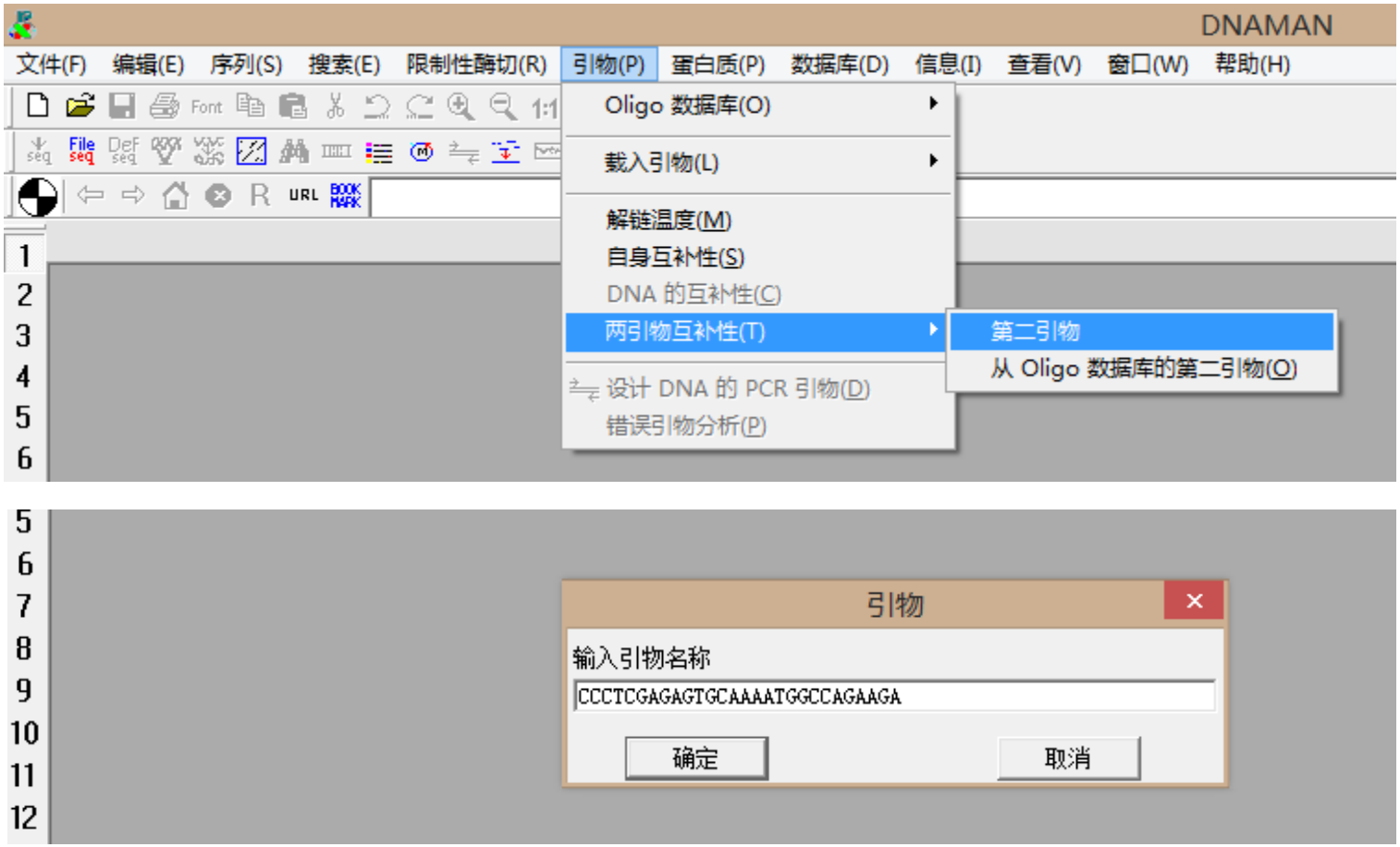


SP-Ha34609F(BamH1):CGGGATCCATGAGTTTATCAATATTTTTGTGG
SP-Ha34609R(Xho1)：CCCTCGAGAGTGCAAAATGGCCAGAAGA
NSP-Ha34609F(BamH1):CGGGATCCATGCATAAATCGTATCAACAACAG
NSP-Ha34609R(Xho1):CCCTCGAGAGTGCAAAATGGCCAGAAGA

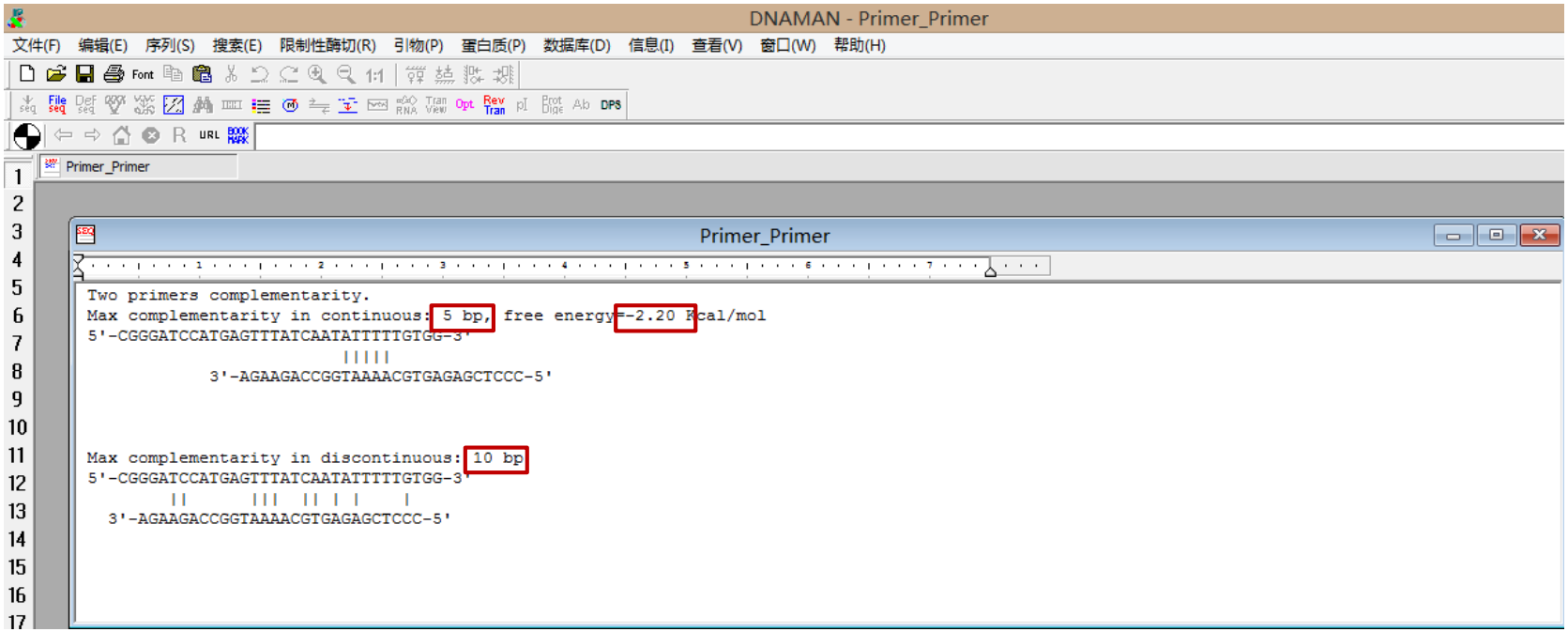
9.用DNAMAN检测引物特异性



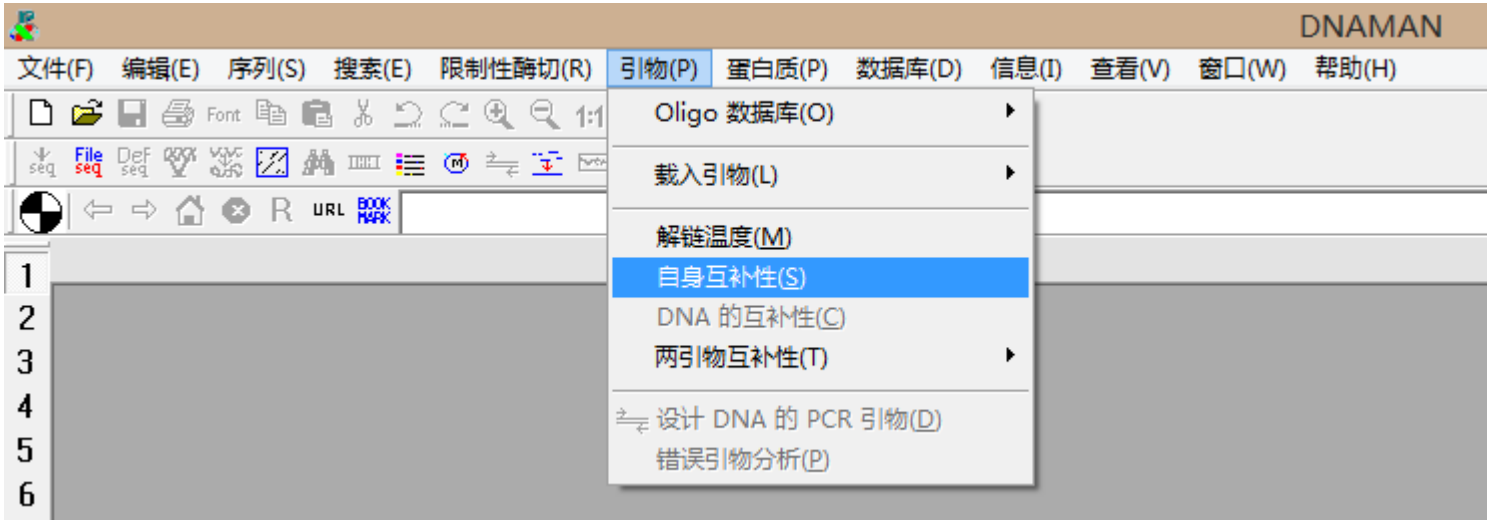
9.用DNAMAN检测引物特异性

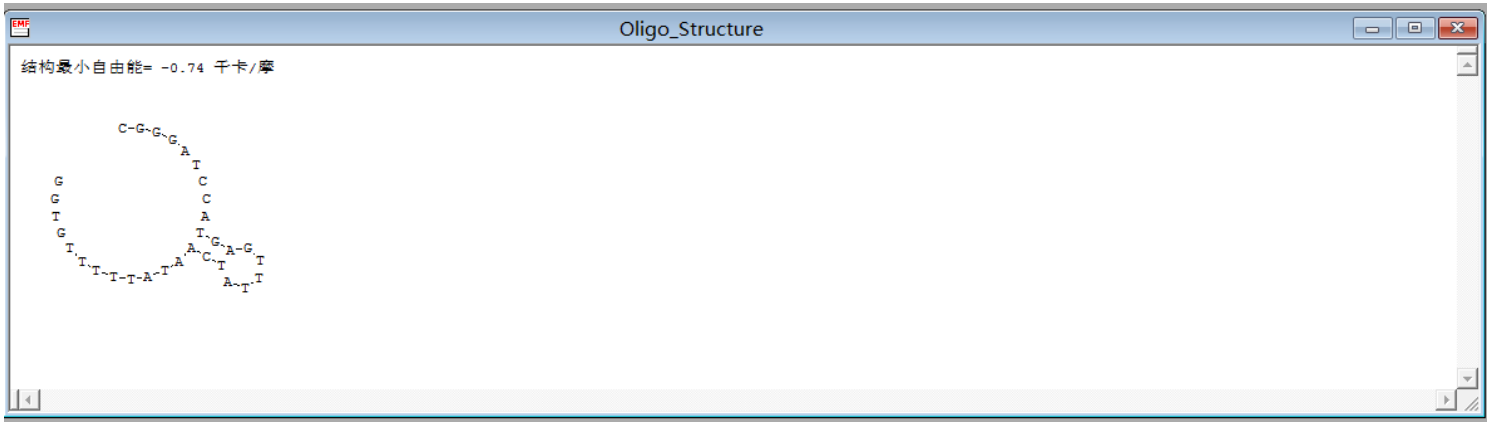


9.用DNAMAN检测引物特异性

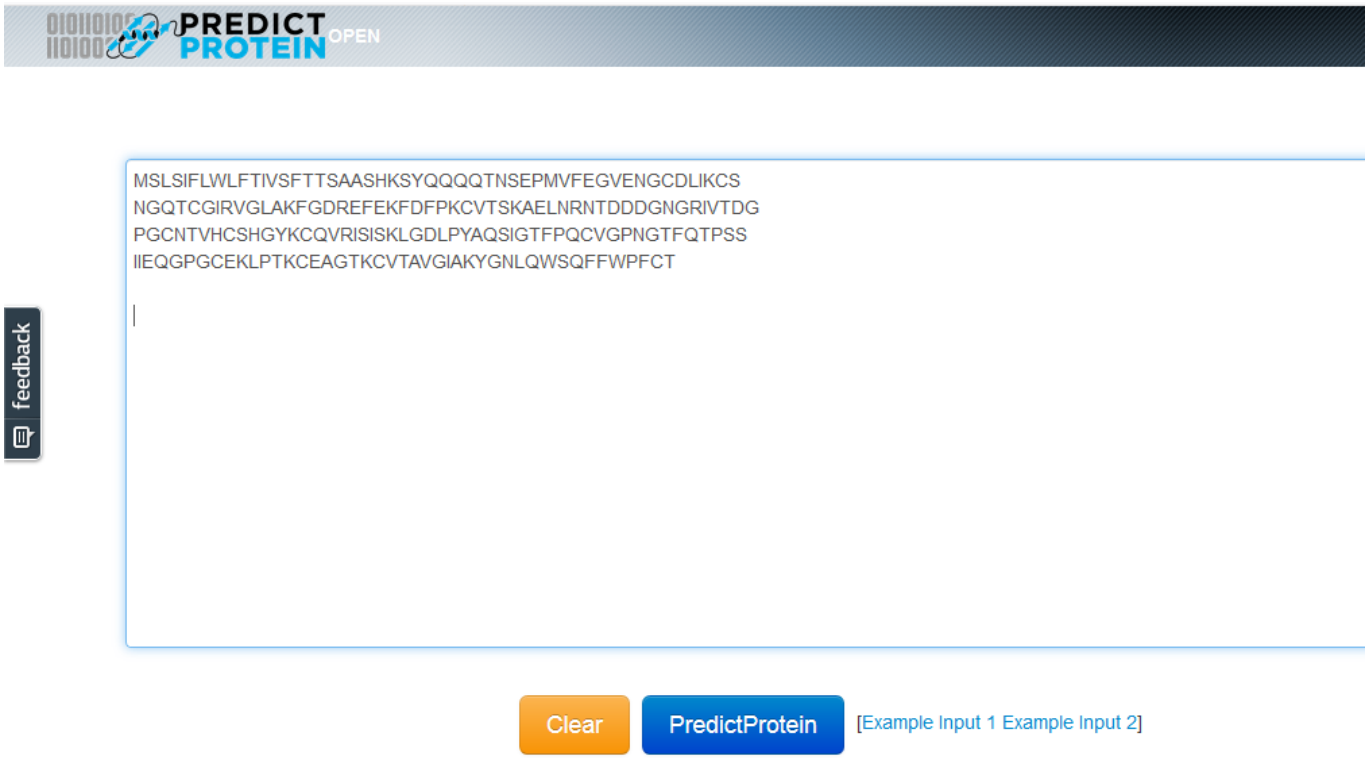


9.用DNAMAN检测引物特异性

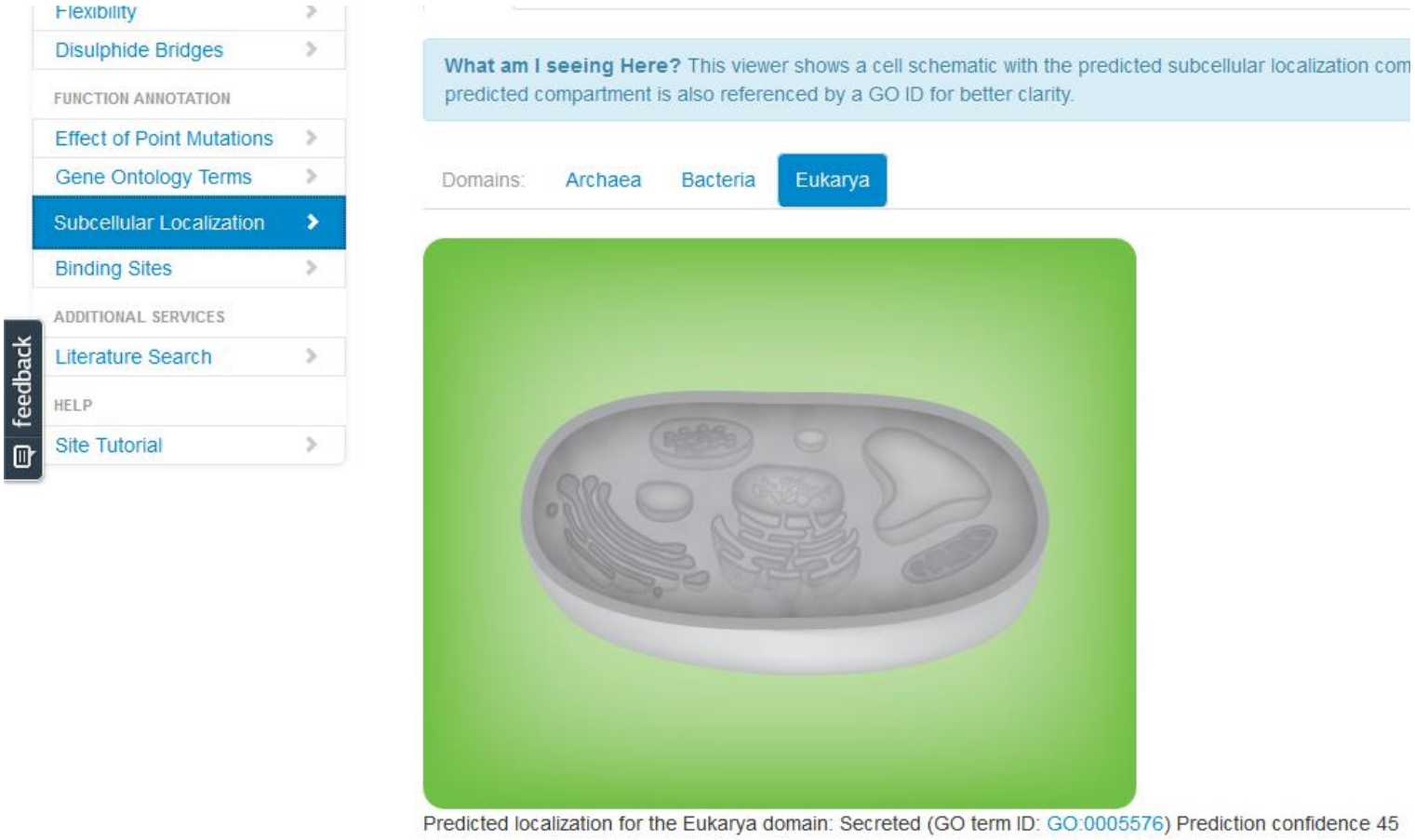




10.亚细胞定位的预测


PREDICT PROTEIN OPEN
feedback
MSLSIFLWLFTIVSFTTSAASHKSYQQQQTNSEPMVFEGVENGCDLIKCS
NGQTCGIRVGLAKFGDREFEKFDFPKCVTSKAELNRNTDDDGNGRIVTDG
PGCNTVHCSHGYKCQVRISISKLGDLPYAQSIGTFPQCVGPNGTFQTPSS
IIEQGPGCEKLPTKCEAGTKCVTAVGIAKYGNLQWSQFFWPFCT
Clear
PredictProtein
[Example Input 1 Example Input 2]

10.亚细胞定位的预测


Flexibility
Disulphide Bridges
FUNCTION ANNOTATION
Effect of Point Mutations
Gene Ontology Terms
Subcellular Localization
Binding Sites
ADDITIONAL SERVICES
Literature Search
HELP
Site Tutorial
feedback
What am I seeing Here? This viewer shows a cell schematic with the predicted subcellular localization com
predicted compartment is also referenced by a GO ID for better clarity.
Domains: Archaea Bacteria Eukarya
Predicted localization for the Eukarya domain: Secreted (GO term ID: GO:0005576) Prediction confidence 45

THANKS

谢谢！