

# Come progettare una proteina di fusione per yeast cell surface display

Per produrre un biocatalizzatore che degrada il PET, il lievito *Pichia pastoris* può essere ingegnerizzato per esporre sulla superficie della cellula l'enzima PETasi del batterio *Ideonella sakaiensis* (*Piscinibacter sakaiensis*) che idrolizza il PET.

La sequenza codificante della PETasi di *I. sakaiensis* deve essere fusa alla sequenza codificante della proteina di ancoraggio GCW51, una proteina della parete cellulare che è covalentemente legata a glucani mediante un'ancora GPI al C-terminale.

1. Progettare la fusione
2. Progettare la strategia di clonaggio
3. Progettare i primer da utilizzare nelle PCR per generare la fusione e clonarla nel vettore di espressione

Il sito di clonaggio multiplo del vettore è mostrato nell'allegato 1

La sequenza codificante della PETasi e di GCW51 è mostrata nell'allegato 2 e 3 insieme alle mappe di restrizione

Il codice genetico è mostrato nell'allegato 4

Progettare la fusione:

- In quale ordine vuoi unire le due sequenze codificanti?
- Aggiungeresti un tag (dove e perchè)?
- Aggiungeresti una sequenza linker?

Progettare la strategia di clonaggio basata sulla PCR:

- Scelta degli enzimi di restrizione
- Clonaggio sequenziale
- Clonaggio simultaneo mediante overlap extension PCR

Progettare i primer

- Porzione che appaia sul template (20-25 basi)
- Estensioni al 5' (lunghezza variabile)

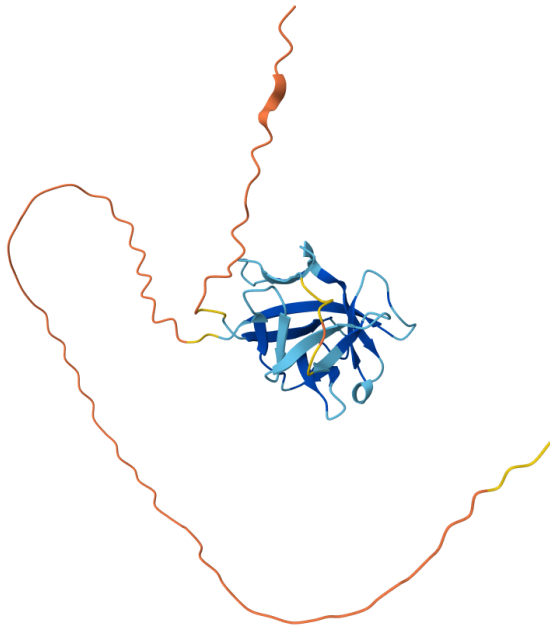
Ricordarsi di verificare che la traduzione della sequenza codificante della proteina di fusione sia corretta!

Lista dei siti da consultare:

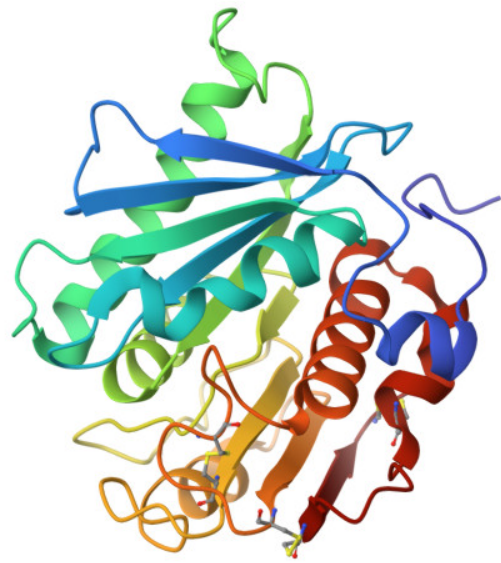
Banca dati di sequenze di proteine UNIPROT <https://www.uniprot.org/>

Banca dati di strutture di proteine PDB <https://www.rcsb.org/>

Grafica molecolare <https://pymol.org/>



GCW51 (AF-C4R7M8-F1)

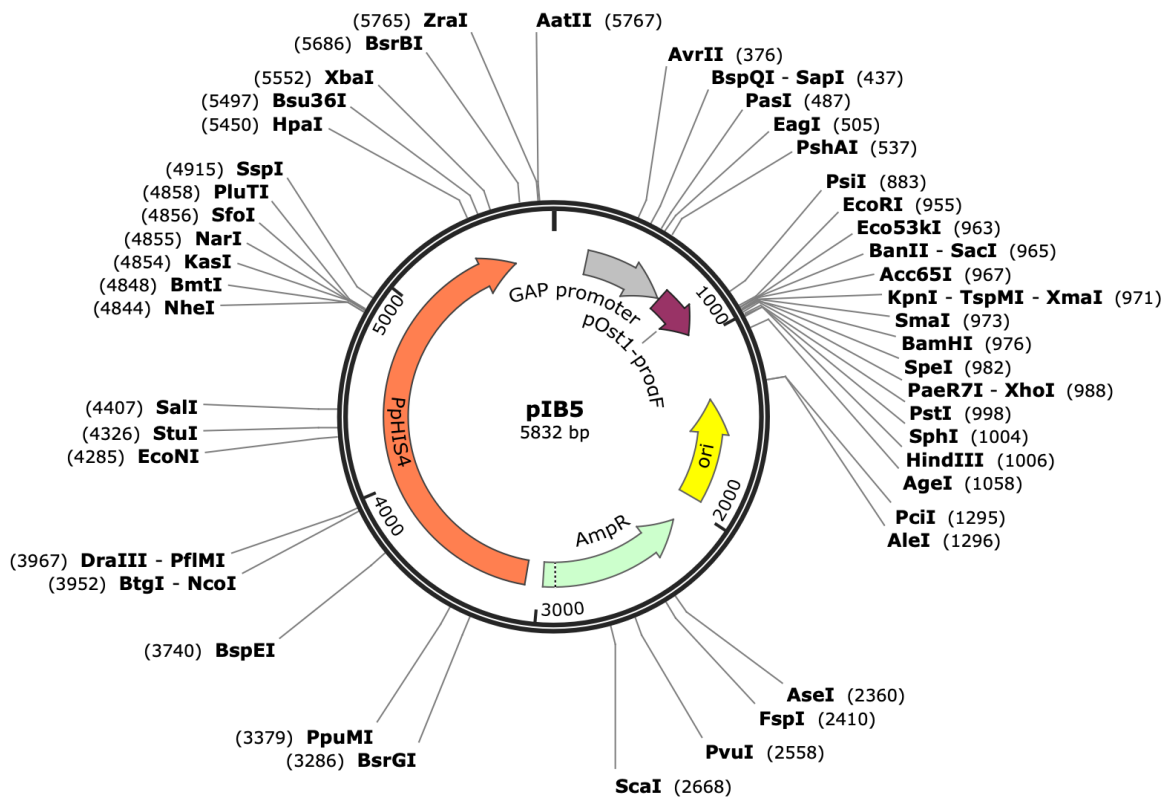


PETase (9LMV)

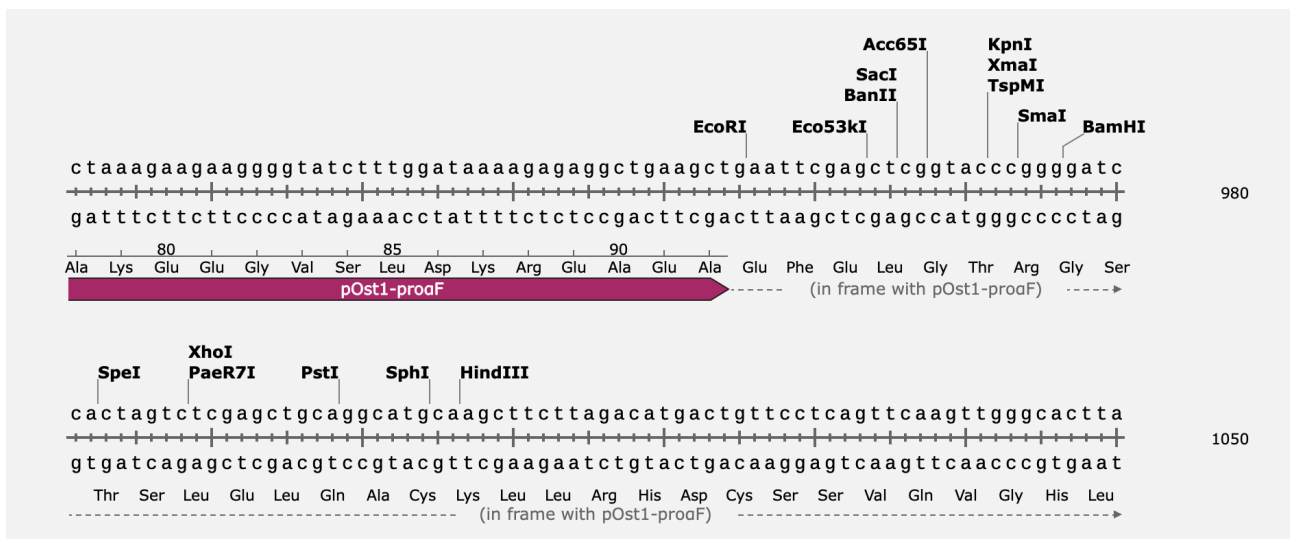
```
>tr|C4R7M8|C4R7M8_KOMPG Uncharacterized protein OS=Komagataella phaffii (strain
GS115 / ATCC 20864) OX=644223 GN=PAS_chr4_0359 PE=4 SV=1
MYRKLLPLALLSVPVFADDDSLPFVIVNSASGETHDGTYWGVNAGAVVPSSTGVRFVV
NDDGELEGNDDEEVEVTSNGFLTLRDDNDENEGFSLTDDDDPPVLLLNQRPTVWICGSDDD
ARIALGSQSPQDDCVEYSIEVQLQSGSRSGSSTRTSSRTTGTSATSATSATSGTSATGTT
TGSTSTATDGAHKLVGGLSGLAGGVAIALLI
```

```
>sp|A0A0K8P6T7|PETH_PISS1 Poly(ethylene terephthalate) hydrolase
OS=Piscinibacter sakaiensis OX=1547922 GN=ISF6_4831 PE=1 SV=1
MNFPRASRLMQAAVLGGLMAVSAATAQTNPYARGPNPTAASLEASAGPFTVRSFTVSRP
SGYGAGTVYYPTNAGGTVGAIIVPGYTARQSSIKWWGPRLASHGFVVITIDNSTLDQP
SSRSSQQMAALRQVASLNGTSSSPIYGVDTARMGVMGWSMGGGSLISAANNPSLKAAA
PQAPWDSSTNFSSVTVPPTLI FACENDSIAPVNSSALPIYDSMSRNAKQFLEINGGSHSCA
NSGNSNQALIGKKGVAWMKRFMDNDTRYSTFACENPNSTRVSDFRTANCS
```

# Allegato 1



## Il sito di clonaggio multiplo di pIB5





SstII AccB1I Hsp92I HaeII BstDSI BstEII  
 NspBII BanI Eco64I Hin1I DsaI Eco91I  
 DsaI SacII BshNI BbiII Bsp143II Eco065I  
 ctgaaagccgcgccgagccccgtgggacagctcgaccaacttctcgtcgggtcaccgtgcccacgctgac base pairs  
 gactttcggcgccggtccgggacacctgtcgagctgggtgaagagcagccagtggcacgggtgagactag 526 to 600  
 BstDSI Sfr303I Msp17I BbeI Eco0109I BstPI  
 MspA1I KasI AcyI NarI BstH2I PspEI  
 KspI Cfr42I BsaHI EheI DraII

HincII NspI  
 Eam1105I  
 AspEI  
 ttgcctgagagaacgacagcatcgccccgtcaactcgtccgcctgcccgatctacgacagcatgctcgcgcaat base pairs  
 aagcggacgctcttctgtcgtgtagcggggccagttgagcaggcgggacggctagatgctgtcgtacagcgcgta 601 to 675  
 HindII AhdI  
 EclHKI

PaeR7I HaeII  
 Eco88I AlwNI  
 BsrDI XmnI BcoI Sfr274I NspBII  
 gcgaagcagttcctcgagatcaacgggtgctcgcactcctgcccacacagcggcaacagcaaccagggcgtgac base pairs  
 cgcttcgtcaagcagctctagttgccaccgagcgtgaggacgcggttgcgcgcttgcgttgggtccgcgactag 676 to 750  
 Asp700I AvaI MspA1I Bsp143II  
 Ama87I BstH2I  
 BsoBI XhoI

Asp700I  
 XmnI BstH2I  
 Aor51HI  
 ggcaagaagggcgtggcctggatgaagcgttcatggacaacgacacgcgctactccaccttcgctgagagaac base pairs  
 ccgttcttcccgcaccggactacttccggaagtacctgttgcgtgctgagatgaggtggaagcggacgctcttg 751 to 825  
 AfeI HaeII  
 Eco47III  
 Bsp143II

MspA1I  
 PstI  
 MslI SfcI PvuII  
 ccgaacagcaccgcgtgctcggacttccgcaccgcgaactgcagctga base pairs  
 ggcttgcgtgggcccacagcctgaaggcgtggcgcttgacgctcgact 826 to 873  
 BstSFI  
 NspBII

**Table by Enzyme Name**

Enzyme name	No. cuts	Positions of sites	Recognition sequence
AccB1I	6	190 196 226 235 415 537	g/gyrcc
AccBSI	2	185 371	gagcgg
AccI	1	446	gt/mkac
AcyI	3	191 236 538	gr/cgyc
AfeI	1	778	agc/gct
AhdI	2	357 663	gacnnn/nngtc
AlwNI	1	744	cagnnn/ctg
Ama87I	2	177 688	c/ycgrg
Aor51HI	1	778	agc/gct
ApaI	1	295	gggcc/c
Asp700I	2	682 778	gaann/nnttc
AspEI	2	357 663	gacnnn/nngtc
AvaI	2	177 688	c/ycgrg
BanI	6	190 196 226 235 415 537	g/gyrcc

BanII	1	295							grgcy/c
BbeI	3	194	239	541					ggcgc/c
BbiII	3	191	236	538					gr/cgyc
BcoI	2	177	688						c/ycgrg
BglI	2	64	182						gccnnnn/nggc
BsaHI	3	191	236	538					gr/cgyc
BsaOI	1	68							cgry/cg
BseI18I	3	139	193	220					r/ccggy
BsePI	1	265							g/cgcgc
Bsh1285I	1	68							cgry/cg
BshNI	6	190	196	226	235	415	537		g/gyrcc
BsiEI	1	68							cgry/cg
BsoBI	2	177	688						c/ycgrg
Bsp120I	1	291							g/ggcc
Bsp143II	6	140	194	239	541	745	780		rgcgc/y
BsrBI	2	185	371						gagcgg
BsrDI	1	676							gcaatg
BsrFI	3	139	193	220					r/ccggy
BssAI	3	139	193	220					r/ccggy
BssHII	1	265							g/cgcgc
BstD102I	2	185	371						gagcgg
BstDSI	4	99	230	533	550				c/crygg
BstEII	1	579							g/gtnacc
BstH2I	6	140	194	239	541	745	780		rgcgc/y
BstMCI	1	68							cgry/cg
BstPI	1	579							g/gtnacc
BstSFI	1	864							c/tryag
BstZI	1	65							c/ggccg
Cfr10I	3	139	193	220					r/ccggy
Cfr42I	2	102	536						ccgc/gg
CfrI	4	56	65	383	509				y/ggcer
DraII	2	291	546						rg/gnccy
DsaI	4	99	230	533	550				c/crygg
EaeI	4	56	65	383	509				y/ggcer
EagI	1	65							c/ggccg
Eam1105I	2	357	663						gacnnn/nngtc
EclHKI	2	357	663						gacnnn/nngtc
EclXI	1	65							c/ggccg
Eco24I	1	295							grgcy/c
Eco47III	1	778							agc/gct
Eco52I	1	65							c/ggccg
Eco64I	6	190	196	226	235	415	537		g/gyrcc
Eco88I	2	177	688						c/ycgrg
Eco91I	1	579							g/gtnacc
EcoNI	1	28							cctnn/nnnagg
EcoO109I	2	291	546						rg/gnccy
EcoO65I	1	579							g/gtnacc
EheI	3	192	237	539					ggc/gcc
FriOI	1	295							grgcy/c
HaeII	6	140	194	239	541	745	780		rgcgc/y
HinI	3	191	236	538					gr/cgyc
HincII	2	447	633						gty/rac
HindII	2	447	633						gty/rac
Hsp92I	3	191	236	538					gr/cgyc
KasI	3	190	235	537					g/gcgc
KspI	2	102	536						ccgc/gg
MroNI	3	139	193	220					g/ccggc
MslI	2	463	838						caynn/nnrtg
Msp17I	3	191	236	538					gr/cgyc
MspAlI	4	101	535	725	869				cmg/ckg
NaeI	3	141	195	222					gcc/ggc
NarI	3	191	236	538					gg/cgcc
NgoAIV	3	139	193	220					g/ccggc

NgoMI	3	139	193	220	g/ccggc	
NspBII	4	101	535	725	869	cmg/ckg
NspI	1	666				rcatg/y
PaeR7I	1	688				c/tcgag
PspEI	1	579				g/gtnacc
PspOMI	1	291				g/ggcc
PstI	1	868				ctgca/g
PvuII	1	869				cag/ctg
SacII	2	102	536			ccgc/gg
SalI	1	445				g/tcgac
SfcI	1	864				c/tryag
Sfii	1	64				ggcnnnn/nggcc
Sfr274I	1	688				c/tcgag
Sfr303I	2	102	536			ccgc/gg
SgrAI	1	220				cr/ccggyg
SstII	2	102	536			ccgc/gg
XhoI	1	688				c/tcgag
XmaIII	1	65				c/ggccg
XmnI	2	682	778			gaann/nnttc

---

The following endonucleases were selected but don't cut this sequence:

AatI, AatII, Accl13I, Accl16I, Acc65I, AccB7I, AccIII, AclNI, AcsI, AflII, AflIII, AgeI, Alw21I, Alw44I, AocI, ApaLI, ApoI, AscI, AseI, AsnI, Asp718I, AspHI, AspI, AtsI, AviII, AvrII, BalI, BamHI, BanIII, BbrPI, BbsI, BbuI, Bbv12I, Bbv16II, BcgI, BclI, BfrI, BglII, BlnI, BlpI, BpiI, BpmI, Bpu1102I, Bpu14I, BpuAI, Bsa29I, BsaAI, BsaBI, BsaI, BsaMI, BsaWI, BscI, Bse21I, Bse8I, BseAI, BseCI, BseRI, BsgI, Bsh1365I, BsiHKAI, BsiI, BsiMI, BsiWI, BsmBI, BsmI, Bsp106I, Bsp119I, Bsp13I, Bsp1407I, Bsp1720I, Bsp19I, Bsp68I, BspCI, BspDI, BspEI, BspHI, BspLU11I, BspMI, BspTI, BspXI, BsrBRI, BsrGI, BssSI, BssT1I, Bst1107I, Bst98I, BstBI, BstI, BstSNI, BstX2I, BstXI, BstYI, Bsu15I, Bsu36I, CciNI, CelII, Cfr9I, ClaI, CpoI, Csp45I, CspI, CvnI, DraI, DraIII, DrdI, Eam1104I, EarI, Ecl136II, Eco105I, Eco130I, Eco147I, Eco255I, Eco31I, Eco32I, Eco57I, Eco72I, Eco81I, EcoICRI, EcoRI, EcoRV, EcoT14I, EcoT22I, ErhI, Esp1396I, Esp3I, FauNDI, FbaI, FseI, FspI, GsuI, HindIII, HpaI, Kpn2I, KpnI, Ksp22I, Ksp632I, LspI, MamI, MfeI, MflI, MluI, MluNI, Mph1103I, MroI, MscI, MspCI, MunI, Mva1269I, NcoI, NdeI, NheI, NotI, NruI, NsiI, NspV, PacI, PaeI, Pfl23II, PflMI, PinAI, Ple19I, PmaCI, Pme55I, PmeI, PmlI, Ppu10I, PpuMI, PshAI, PshBI, Psp124BI, Psp1406I, Psp5II, PspAI, PspALI, PspLI, PstNHI, PvuI, RcaI, RsrII, SacI, SapI, SbfI, ScaI, SexAI, SfuI, SgfI, SmaI, SmiI, SnaBI, SpeI, SphI, SplI, SrfI, Sse8387I, SseBI, SspBI, SspI, SstI, StuI, StyI, SunI, SwaI, Tth111I, Van91I, Vha464I, VneI, VspI, XbaI, XcmI, XhoII, XmaI, Zsp2I

## Allegato 3

### GCW51 636 base pairs

Bbv16II  
BpiI  
atgtatagaaaattgcttcccttgctttgctcagtggtccagctcttcgccgatgacgatgactcattacctttt base pairs  
tacatatcttttaacgaaggggaacgaaacgagtcacaaggtcagaagcgggctactgctactgagtaatggaaaa 1 to 75  
BpuAI  
BbsI

HindII  
HpaI  
gttattgttaactccgctcaggagaaacctatgatggaacttactgggggtggaataacgccggagcagtcggt base pairs  
caataacaattgagggcgcagtcctctttgggtactaccttgaatgacccacacttattgcggcctcgtcagcaa 76 to 150  
HincII

BpmI  
MslI HincII  
ccaagctccactggagtgcgatttggtggtcaacgacgacgggtgagttggaaggtaatgacgaggaagttgaagtg base pairs  
ggttcgaggtgacctcacgctaaacaccagttgctgctgccactcaaccttcattactgctcctcaacttcac 151 to 225  
GsuI HindII

Bst98I Bsh1365I  
Vha464I Bse8I  
MspCI BsaBI  
acctcaaacgggtttttgaccttaagagatgacaatgatgaaaatgaggggttttctctcacagatgatgatcca base pairs  
tggagtttgcccaaaaactggaattctctactgttactacttttactcccaaaaagagagtgctactactaggt 226 to 300  
BspTI BsrBRI  
BfrI MamI  
AflII

PinAI HinII  
AgeI Cfr10I BstX2I BsaHI  
Bse118I MflI Msp17I  
ccggttttgctgttaaatcgtaaaactcctaccgtggtgatatgtggatctgatgatgacgcccgtattgctctg base pairs  
ggccaaaacgacaatttagcagtttgaggatggcacacctatacacctagactactactgcgggcataacgagac 301 to 375  
BsaWI BstYI BbiII  
BssAI XhoII AcyI  
BsrFI Hsp92I

Eco255I NspBII  
Acc113I MspA1I  
ggttcacaatcgccacaagatgattgtgtagagtactccattgaagttcagctgcaaaagtggttcaagaagcggg base pairs  
ccaagtgttagcgggtgttctactaacacatctcatgaggttaacttcaagtcgacgtttcaccaagttcttcgccc 376 to 450  
ScaI PvuII

tccagcacaagaacaagcagtagaacaactggaaccagtgcaaccagtgcaaccagtgcaaccagtggaaccagt base pairs  
aggctggttcttctgttcgtcatcttgttgaccttggtcacggttggtcacggttggtcaccttggtca 451 to 525

Eco24I  
AlwNI BanII XcmI  
gcaacagggacgactactggaagcacctcgacagctactgatggagcccacaagcttggggcggctgtctgga base pairs  
cgttgtccctgctgatgaccttcgtggagctgtcgtatgactacctcgggtgtttcgaaacaccgcccagacacct 526 to 600  
FriOI  
HindIII

BsrDI  
ttggctgggtggtgttgccattgcctattgatctag base pairs  
aaccgaccaccacaacggtaacgggataactagatc 601 to 636

---

**Table by Enzyme Name**

Enzyme name	No. cuts	Positions of sites	Recognition sequence
Acc113I	1	409	agt/act
AcyI	1	359	gr/cgyc
AflII	1	246	c/ttaag
AgeI	1	300	a/ccggt
AlwNI	1	562	cagnnn/ctg
BanII	1	573	grgcy/c
BbiII	1	359	gr/cgyc
BbsI	1	48	gaagac
Bbv16II	1	48	gaagac
BfrI	1	246	c/ttaag
BpiI	1	48	gaagac
BpmI	1	166	ctggag
BpuAI	1	48	gaagac
BsaBI	1	293	gatnn/nnatc
BsaHI	1	359	gr/cgyc
BsaWI	1	300	w/ccggw
Bse118I	1	300	r/ccggy
Bse8I	1	293	gatnn/nnatc
Bsh1365I	1	293	gatnn/nnatc
BspTI	1	246	c/ttaag
BsrBRI	1	293	gatnn/nnatc
BsrDI	1	623	gcaatg
BsrFI	1	300	r/ccggy
BssAI	1	300	r/ccggy
Bst98I	1	246	c/ttaag
BstX2I	1	346	r/gatcy
BstYI	1	346	r/gatcy
Cfr10I	1	300	r/ccggy
Eco24I	1	573	grgcy/c
Eco255I	1	409	agt/act
FriOI	1	573	grgcy/c
GsuI	1	166	ctggag
HinII	1	359	gr/cgyc
HincII	2	84 180	gty/rac
HindII	2	84 180	gty/rac
HindIII	1	577	a/agctt
HpaI	1	84	gtt/aac
Hsp92I	1	359	gr/cgyc
MamI	1	293	gatnn/nnatc
MflI	1	346	r/gatcy
MslI	1	163	caynn/nrntg
Msp17I	1	359	gr/cgyc
MspA1I	1	426	cmg/ckg
MspCI	1	246	c/ttaag
NspBII	1	426	cmg/ckg
PinAI	1	300	a/ccggt
PvuII	1	426	cag/ctg
ScaI	1	409	agt/act
Vha464I	1	246	c/ttaag
XcmI	1	580	ccannnnn/nnntgg
XhoII	1	346	r/gatcy

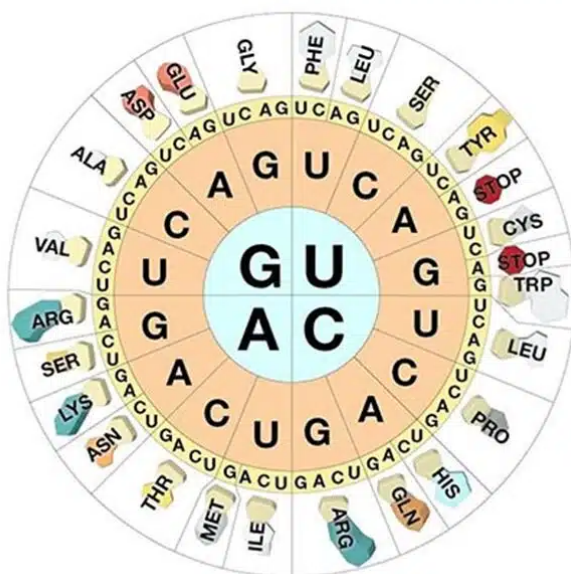
---

The following endonucleases were selected but don't cut this sequence:

AatI, AatII, AccI6I, Acc65I, AccB1I, AccB7I, AccBSI, AccI, AccIII, AclNI, AcsI, AfeI, AflIII, AhdI, Alw21I, Alw44I, Ama87I, AocI, Aor51HI, ApaI, ApaLI, ApoI, AscI, AseI, AsnI, Asp700I, Asp718I, AspEI, AspHI, AspI, AtsI, AvaI, AviII, AvrII, BalI, **BamHI**, BanI, BanIII, BbeI, BbrPI, BbuI, Bbv12I, BcgI, BclI, BcoI, BglI, BglII, BlnI, BlpI, Bpu1102I, Bpu14I, Bsa29I, BsaAI, BsaI, BsaMI, BsaOI, BscI, Bse21I, BseAI, BseCI, BsePI, BseRI, BsgI, Bsh1285I, BshNI, BsiEI, BsiHKAI, BsiI, BsiMI, BsiWI, BsmBI, BsmI, BsoBI, Bsp106I, Bsp119I, Bsp120I, Bsp13I, Bsp1407I, Bsp143II, Bsp1720I, Bsp19I, Bsp68I, BspCI, BspDI, BspEI, BspHI, BspLU11I, BspMI, BspXI, BsrBI, BsrGI, BssHII, BssSI, BssTII, Bst1107I, BstBI, BstD102I, BstDSI, BstEII, BstH2I, BstI, BstMCI, BstPI, BstSFI, BstSNI, BstXI, BstZI, Bsu15I, Bsu36I, CciNI, CelII, Cfr42I, Cfr9I, CfrI, ClaI, CpoI, Csp45I, CspI, CvnI, DraI, DraII, DraIII, DrdI, DsaI, EaeI, EagI, Eam1104I, Eam1105I, EarI, Ecl136II, EclHKI, EclXI, Eco105I, Eco130I, Eco147I, Eco31I, Eco32I, Eco47III, Eco52I, Eco57I, Eco64I, Eco72I, Eco81I, Eco88I, Eco91I, EcoICRI, EcoNI, EcoO109I, EcoO65I, **EcoRI**, EcoRV, EcoT14I, EcoT22I, EheI, ErhI, Esp1396I, Esp3I, FauNDI, FbaI, FseI, FspI, HaeII, KasI, Kpn2I, KpnI, Ksp22I, Ksp632I, KspI, LspI, MfeI, MluI, MluNI, Mph1103I, MroI, MroNI, MscI, MunI, Mva1269I, NaeI, NarI, NcoI, NdeI, NgoAIV, NgoMI, NheI, NotI, NruI, NsiI, NspI, NspV, PacI, PaeI, Paer7I, Pfl23II, PflMI, Ple19I, PmaCI, Pme55I, PmeI, PmlI, Ppu10I, PpuMI, PshAI, PshBI, Psp124BI, Psp1406I, Psp5II, PspAI, PspALI, PspEI, PspLI, PspOMI, **PstI**, PstNHI, PvuI, RcaI, RsrII, **SacI**, SacII, SalI, SapI, SbfI, SexAI, SfcI, SfiI, Sfr274I, Sfr303I, SfuI, SgfI, SgrAI, **SmaI**, SmiI, SnaBI, SpeI, SphI, SplI, SrfI, Sse8387I, SseBI, SspBI, SspI, SstI, SstII, StuI, StyI, SunI, SwaI, Tth111I, Van91I, VneI, VspI, XbaI, **XhoI**, XmaI, XmaIII, XmnI, Zsp2I

## Allegato 4

# The Genetic Code



		Second Letter					
		T	C	A	G		
First Letter	T	TTT } Phe TTC } TTA } Leu TTG }	TCT } TCC } Ser TCA } TCG }	TAT } Tyr TAC } TAA } Stop TAG } Stop	TGT } Cys TGC } TGA } Stop TGG } Trp	T	C
	C	CTT } CTC } Leu CTA } CTG }	CCT } CCC } Pro CCA } CCG }	CAT } His CAC } CAA } Gln CAG }	CGT } CGC } Arg CGA } CGG }	C	A
	A	ATT } ATC } Ile ATA } ATG } Met	ACT } ACC } Thr ACA } ACG }	AAT } Asn AAC } AAA } Lys AAG }	AGT } Ser AGC } AGA } Arg AGG }	A	G
	G	GTT } GTC } Val GTA } GTG }	GCT } GCC } Ala GCA } GCG }	GAT } Asp GAC } GAA } Glu GAG }	GGT } GGC } Gly GGA } GGG }	G	
							Third Letter