

# Quina és la funció del gen clonat?

Margarita Soriano, IES Verdaguer, Barcelona ([bacillussubti@yahoo.es](mailto:bacillussubti@yahoo.es))

Francisca Guerola, IES Frederic Mompou, St. Vicenç dels Horts ([mguerola@xtec.net](mailto:mguerola@xtec.net))

*Presentem una activitat didàctica complementària per a alumnes de segon de batxillerat. A partir d'una seqüència de nucleòtids del genoma de Paenibacillus barcinonensis determinem la regió codificant d'un gen, la proteïna corresponent i la seva funció. Com a annex s'inclou el guió per a l'alumnat.*

Per a l'alumnat de Batxillerat no és fàcil entendre el desenvolupament i l'aplicació de la genètica molecular i la enginyeria genètica, ja que són temes molt nous que estan evolucionant constantment. Una part fonamental de la nostra feina és ajudar-los a aconseguir-ho.

En aquesta activitat els alumnes posen en pràctica conceptes estudiats a diferents unitats didàctiques estudiades anteriorment: Proteïnes, Enzims, Microbiologia, Genètica molecular i Enginyeria genètica.

Al proposar aquesta pràctica pretenem:

- Acostar als alumnes el treball científic real de un laboratori modern de genètica.
- Fer un repàs de la biologia de segon de batxillerat.
- Veure la relació entre la seqüenciació de genomes i la traducció a proteïnes, fent-los conscients de la dificultat i meticulositat d'aquest treball.

Al final l'alumne ha de produir un text explicatiu de tots els processos que ha seguit en l'activitat.

## INTRODUCCIÓ

L'objectiu global de la Genètica és l'anàlisi de l'estructura i funció dels genomes i dels gens.

Les primeres evidències sobre la funció dels gens van sorgir de l'observació de correlacions entre determinades mutacions i la deficiència d'un enzim o d'altre tipus de proteïnes. La comprensió de l'estructura i funció dels gens va experimentar un gran avenç quan es va establir la relació entre la presència de mutacions en un gen i la presència de

canvis d'aminoàcids específics en la proteïna corresponent. Aquests coneixements, junt amb el descobriment de la naturalesa del DNA i del codi genètic, van ser molt importants per posar de manifest la naturalesa bàsica del gen.

Malgrat això, abans de l'últim terç del segle XX el seu estudi havia estat només per mitjans indirectes, mai s'havia aïllat un gen ni analitzat la seva seqüència. Encara que s'havia aïllat DNA a partir de teixit viu, amb una consistència de massa molt viscosa, com podria aïllar-se un gen específic entre aquesta "massa de fills embolicats"?

L'enginyeria genètica ho ha fet possible, i en l'actualitat s'aïllen gens i altres parts del genoma rutinàriament.

L'aïllament de gens ens permet:

- la determinació de la seqüència nucleotídica i a partir d'aquesta informació es poden establir les característiques del gen
- la comparació entre les seqüències de DNA de diferents gens pot proporcionar informació sobre la seva evolució
- la determinació de la seqüència d'aminoàcids mitjançant la conversió de la seqüència de nucleòtids dictada pel codi genètic, i a partir d'aquesta conèixer la funció de la proteïna
- l'estudi de la modificació de la seqüència de nucleòtids per determinar la funció de cada aminoàcid en la proteïna
- la generació d'organismes transgènics (organismes amb gens procedents d'altres organismes), per la utilització en la investigació tant bàsica com aplicada

Avui dia la determinació de la funció d'una proteïna s'ha facilitat molt degut a la creació i publicació de bases de dades d'accés públic. Una de les més importants i utilitzades és *SwissProt*, a la qual es pot accedir a través del sistema *ExpPASy* (fig. 1) de l'Institut Suís de Bioinformàtica.

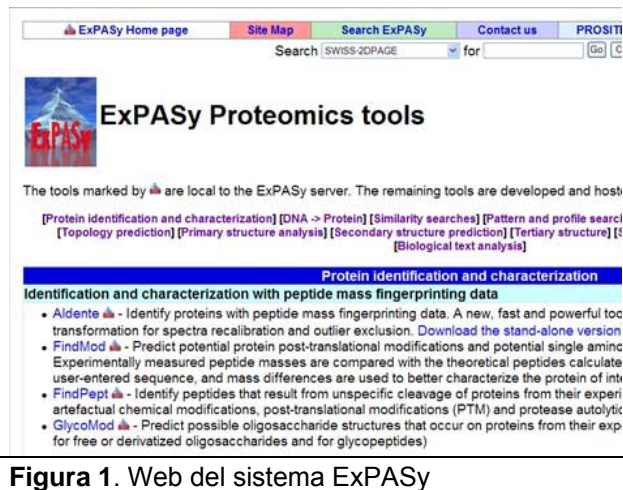


Figura 1. Web del sistema ExpPASy

Cada seqüència de proteïnes és una fitxa de *SwissProt*, en la qual, a més de la estructura primària de la proteïna, també ens ofereix altres tipus d'informacions organitzades en diferents camps. Al voltant de *SwissProt* s'han desenvolupat altres bases de dades i utilitats per a extraure la màxima informació possible a partir de la estructura primària de les proteïnes.

Els diferents exercicis que proposem permetran descobrir *SwissProt* i altres bases de dades i utilitats.

Després de la seqüenciació completa del genoma humà l'abril del 2003, les bases de dades que organitzen tota la informació del genoma humà estan adquirint cada vegada més importància. Una d'elles és la anomenada *GeneCard*. En aquest tipus de bases de dades podem trobar, entre altres coses, la localització cromosòmica de qualsevol gen humà, variants polimòrfiques o SNPs (*Single Nucleotide Polymorphism*), gens homòlegs en altres eucariotes, etc. Per això és important que l'alumnat tingui contacte amb aquest tipus d'informació.

## Identificació i anàlisi d'una seqüència de DNA

La majoria de seqüències de proteïnes que s'incorporen a les bases de dades provenen de la traducció de seqüències de DNA procedents de projectes genoma. Per aquest motiu s'han desen-

volupat una sèrie de tècniques bioinformàtiques per a localitzar les regions codificants i predir la possible funció de la proteïna codificada, així com per a localitzar en el DNA regions reguladores, zones repetitives, etc.

A l'hora de seqüenciar el DNA hem de tenir en compte si es tracta d'un organisme procariota (només té introns) o eucariota (té introns i exons).

La caracterització d'un fragment de DNA genòmic d'un organisme eucariota és més complicada. En aquest cas és necessari predir els llocs exactes on comencen i acaben tant els introns com els exons. Encara que per a les prediccions disposem de guies, com per exemple els llocs consens de *splicing*, sempre hi ha una certa incertesa amb els resultats. A més, en els gens eucariòtics es dona el fenomen de *splicing alternatiu*, el qual dificulta més la definició del lloc de *splicing*. D'aquesta manera, no sempre un gen codifica una única proteïna. És per això que la pràctica que us proposem parteix del genoma d'un organisme procariota. Creiem que per l'alumnat de batxillerat i per als objectius que volem aconseguir és suficient.

Hem de recordar a l'alumnat que quan es tradueix un fragment de DNA és important definir quina serà la pauta de lectura o *open reading frame* (ORF) que s'utilitzarà. Recordem que hi ha sis pautes de lectura diferents (tres per a una cadena i tres més per a la cadena complementària) per a traduir un fragment de DNA, segons que es comenci per la primera, segona o tercera base del fragment de DNA i el bri de DNA que és transcrit en viu.

És per això que els traductors informàtics donen 6 possibles pautes de lectura, 3 *forward* i 3 *reverse*. D'aquestes 6 possibilitats normalment s'escull la pauta de lectura més probable: llavors es diu que s'ha fet una *conceptual translation*, és a dir, s'ha triat la regió codificant més probable.

Una vegada coneguda l'estructura primària més probable de la proteïna es pot predir quina és la funció que té. Per fer-ho es busca en les diferents bases de dades a quines proteïnes ja seqüenciades s'assembla la nostra seqüència problema. Si hi ha un nivell de similitud elevat al llarg de tota la proteïna i només es troben similituds amb una única funció, es pot atribuir aquesta funció a la proteïna problema.

## Objectius

El que es pretén en aquesta activitat està ben relacionat amb els objectius de la matèria de Biologia de Batxillerat:

- Conèixer les tècniques de seqüenciació d'un gen
- Entendre els mecanismes de duplicació, transcripció i traducció (objectiu terminal 37)
- Investigar l'existència de pautes de lectures (ORFs) en una seqüència de DNA (procediment 2.1)
- Identificar la funció de seqüències problema de DNA i proteïnes (Investigar la funció d'un gen)
- Evidenciar la relació existent entre diferents temes (Estructura de les Macromolècules, Proteïnes, Enzimologia, Genètica Molecular, Enginyeria Genètica, Biologia Cel·lular, Microbiologia)
- Orientar-se en l'embolic de les xarxes de bases de dades de seqüències (procediment 2.1)
- Accedir a recursos genòmics, proteòmics i bibliogràfics
- Aprendre a organitzar la informació obtinguda (procediment 3.1)
- Redacció de la memòria de forma rigorosa i concreta, en estil científic (procediment 4)

## PRÀCTICA I COMENTARIS PER AL PROFESSORAT

### Objectiu

En la següent pràctica volem replicar, transcriure i traduir un gen clonat en un plasmidi. D'aquesta manera podrem conèixer la seqüència d'aminoàcids i a partir d'ella identificar la seva funció.

A partir del fragment de DNA seqüenciat, inserit en el plasmidi, hem de:

- Replicar el fragment de DNA ([objectiu terminal 37](#)).
- Transcriure la seqüència de nucleòtids ([objectiu terminal 37](#))
- i posteriorment traduir-la ([objectius terminals 37 i 38](#)).
- saber quina és la part del fragment de DNA que correspon al gen que volem investigar ([objectiu terminal 36](#))
- i conèixer la funció de la proteïna ([objectius terminals 8 i 33](#)).

Cal tenir en compte que el DNA bacterià no té introns. Això significa que cada gen codifica per a una proteïna.

### Continguts de la pràctica

#### Fets:

5.7 Replicació i expressió del material genètic.

#### Procediments:

- 1 Realització i valoració d'experiències
- 2.1 Utilització de material telemàtic
- 3.1 Elaboració de protocols de pràctiques
- 3.5 Formulació de conclusions
- 4.1 Expressió per escrit, de la informació obtinguda a partir de l'anàlisi dels processos de transcripció i traducció.
- 4.2 Utilització del llenguatge científic.
- 4.3 Elaboració i presentació de treballs
- 5 Aplicació de conceptes estudiats a situacions reals

#### Valors:

- 1 Valoració crítica de la Biologia, així com de les seves aplicacions.
- 3 Interès per les relacions que existeixen entre la Biologia, la societat i la tecnologia.

### Materials

Les bases de dades i eines que necessitaràs per a resoldre l'activitat les pots trobar al full de pràctiques i al [Protein Analysis System \(ExPASy: http://www.expasy.ch/tools/\)](http://www.expasy.ch/tools/). Aquesta serà la pàgina a què pots dirigir-te per a triar l'eina adequada en cada ocasió i respondre les preguntes plantejades.

A més, has de tenir en compte que:

- Al llarg de la sessió s'han de guardar les adreces utilitzades per posar-les com a bibliografia.
- Cal guardar tota la informació que es trobi en un disc o a la memòria USB per a poder respondre així a les preguntes i presentar un informe final de les pràctiques. No s'ha de deixar res en el disc dur de l'ordinador ja que l'endemà podria no estar allí.

### Mètode

A partir d'una seqüència de nucleòtids hem de trobar el gen que volem investigar.

```
tcaggctcgtacaaggggtgagcccgcagggtgctgtagaca
gccagcggctgtttatcaaatctttaaataagggaggatgttgat
gatgaaaaaatgttaacgctcttctgctgctgccggtctgttctcc
atatttggtgtaatgctgcccggctgcgccaaccgctgcaattcaa
cgatcgtgtaccgaagggcagcacctatgatggacaggggaaaa
cctttgtagcgaatccctctacattaggtgacggttctcaagcggagaa
tcagaagcctgtattccggctggaagcaggggcaactttgaaaaatg
ttatttggtgcacccgcagcggatggtgtgattgttaccggcagctgt
aatattccaacggtgtatggaagatgtaggcgaggatgcattgaca
ctgaaatcatcaggaaccggttaattaccggcggagcagcgtataa
```

ggcttacgataaagtgttcagatgaacgctcgggtacgattaatatac  
 aaaaactcctgctcggatgatcggcaagctggtgcgcaaatg  
 gaggaacctcctatgctgtaaacatgacactggataattccaacat  
 ccaacgtgaaggattccattatgctgacggacagcagtgatctcaa  
 gggaagatcacgaacacagttactccaaagtccaacgctgttca  
 agggcttctcaggcaagacgagccagtcggaaatacacagta  
 ttaacgcacgcttcagtaaaaaataaagaagcggctgaagtcaa  
 gtaccacctgctcagccgttgaggagaaccaataaaaacgggatg  
 aaccagtaacatggcctcatcccgttttgattggttggttacagc  
 catgaatctgtatccatcgaataactctctccaggcacctcaacga  
 acagttgcttgggatggagcagcagaaccgctatccaaggagcc  
 ggattgtcatgatctcataagcggttctccagcatgcgagaatatt  
 gaaggataaccagcgttttccggctacagatggacgggttctgcctc  
 ggcaacgggtttgcttgataaagaacaggtcaccaatggttccgg

cttgcttgcagattggcttgcatttcaacagtgatttccgttctccatc  
 gtataataactgaccaga

**1. Abans de ficar-nos en feina practicarem la replicació, transcripció i traducció a partir d'una petita seqüència de nucleòtids.**

Copia la seqüència de nucleòtids senyalada en negra. Fes la cadena complementària d'aquesta seqüència. A continuació transcriu i tradueix la seqüència (fig. 2). El resultat és a la Figura 3.

		Segona lletra				U	C	A	G	Tercera lletra	
		U	C	A	G						
Primera lletra	U	UUU	UCU	UAU	UGU	Fen	Ser	Tir	Gln	U	
		UUC	UCC	UAC	UGC						C
		UUA	UCA	UAA	UGA						A
		UUG	UCG	UAG	UGG						G
	C	CUU	CCU	CAU	CGU	Leu	Pro	His	Arg	U	
		CUC	CCC	CAC	CGC						C
		CUA	CCA	CAA	CGA						A
		CUG	CCG	CAG	CGG						G
	A	AUU	ACU	AAU	AGU	Ile	Tre	Asn	Ser	U	
		AUC	ACC	AAC	AGC						C
		AUA	ACA	AAA	AGA						A
		AUG	ACG	AAG	AGG						G
	G	GUU	GCU	GAU	GGU	Val	Ala	Asp	Gli	U	
		GUC	GCC	GAC	GGC						C
		GUA	GCA	GAA	GGA						A
		GUG	GCG	GAG	GGG						G

**Figura 2.** Pauta per a la lectura de seqüències de mRNA.

<b>1</b>	<b>TAA</b>	<b>GGG</b>	<b>AGG</b>	<b>ATG</b>	<b>TTG</b>	<b>ATG</b>	<b>ATG</b>	<b>AAA</b>	<b>AAA</b>	<b>ATG</b>	<b>TTA</b>	<b>ACG</b>	<b>CTC</b>	<b>TTG</b>	<b>CTG</b>	
<b>DNA</b>	ATT	CCC	TCC	TAC	AAC	TAC	TAC	TTT	TTT	TAC	AAT	TGC	GAG	AAC	GAC	
<b>mRNA</b>	UAA	GGG	AGG	AUG	UUG	AUG	AUG	AAA	AAA	AUG	UUA	ACG	CUC	UUG	CUG	
<b>Aa</b>	Stop	Gly	Arg	Met	Leu	Met	Met	Lys	Lys	Met	Leu	Thr	Leu	Leu	Leu	
<b>2</b>	<b>T</b>	<b>AAG</b>	<b>GGA</b>	<b>GGA</b>	<b>TGT</b>	<b>TGA</b>	<b>TGA</b>	<b>TGA</b>	<b>AAA</b>	<b>AAA</b>	<b>TGT</b>	<b>TAA</b>	<b>CGC</b>	<b>TCT</b>	<b>TGC</b>	<b>TG</b>
<b>DNA</b>	A	TTC	CCT	CCT	ACA	ACT	ACT	ACT	TTT	TTT	ACA	ATT	GCG	AGA	ACG	AC
<b>mRNA</b>	U	AAG	GGA	GGA	UGU	UGA	UGA	UGA	AAA	AAA	UGU	UAA	CGC	UCU	UGC	UG
<b>Aa</b>		Lys	Gly	Gly	Cys	Stop	Stop	Met	Lys	Lys	Cys	Stop	Arg	Ser	Cys	
<b>3</b>	<b>TA</b>	<b>AGG</b>	<b>GAG</b>	<b>GAT</b>	<b>GTT</b>	<b>GAT</b>	<b>GAT</b>	<b>GAA</b>	<b>AAA</b>	<b>AAT</b>	<b>GTT</b>	<b>AAC</b>	<b>GCT</b>	<b>CTT</b>	<b>GCT</b>	<b>G</b>
<b>DNA</b>	AT	TCC	CTC	CTA	CAA	CTA	CTA	CTT	TTT	TTA	CAA	TTG	CGA	GAA	CGA	C
<b>mRNA</b>	UA	AGG	GAG	GAU	GUU	GAU	GAU	GAA	AAA	AAU	GUU	AAC	GCU	CUU	GCU	G
<b>Aa</b>		Ser	Glu	Asp	Val	Asp	Asp	Glu	Lys	Asn	Val	Asn	Ala	Leu	Ala	

**Figura 3.** Els tres possibles resultats.

- a) De la seqüència estudiada on podria començar una hipotètica proteïna?

Els alumnes han de saber que el primer triplet que es tradueix és, normalment, l'AUG, que correspon a la metionina, per tant, la proteïna ha de començar per aquest aminoàcid. En aquest cas hi ha tres metionines juntes. Podria ser qualsevol de les tres. Per precisar més hauríem de conèixer la distància a senyals consens. Són seqüències que es troben a una distància determinada dels codons d'iniciació, i que s'aparellen amb l'extrem 3' del rRNA 16S del ribosoma. En procariontes són semblants a TTGACA (a una distància de 35 nucleòtids de AUG) i a TATAAT (a una distància de 10 nucleòtids del codó AUG). Es mostren assenyalats en blau en la seqüència.

El DNA bacterià presenta una estructura circular de doble cadena. Això significa que el nostre gen pot estar en qualsevol de les dues cadenes. Com les seqüències poden ser molt llargues, podem tenir la seqüència del gen en la cadena original o en la complementària, i a la vegada cadascuna pot tenir tres pautes de lectures o ORF (*open reading frame*), en els laboratoris es treballa amb programes que ens transcriuen i tradueixen la seqüència de la hipotètica proteïna.

- b) A l'exercici a) has traduït la seqüència de mRNA utilitzant una pauta de lectura. A continuació, torna a traduir la seqüència de mRNA utilitzant les altres 2 pautes de lectura. La seqüència d'aminoàcids és la mateixa?

Aquest exercici s'aconsella que es faci en grups de dos alumnes, així podran comparar el resultat de les seqüències d'aminoàcids (fig. 3) i comentar-les. Posteriorment, es farà una posada en comú del grup classe.

**2. A partir de la seqüència de nucleòtids completa donada al principi de la pràctica, la replicarem, transcriurem i traduirem amb un programa informàtic per trobar la seqüència de la hipotètica proteïna:**

- Primer, anirem a la següent pàgina: <http://www.expasy.ch/tools/>
- Clicarem sobre la següent opció:  
*DNA → Protein*
- A continuació farem dos clics sobre: [Translate](#)
- Enganxa la seqüència de nucleòtids en el quadre en blanc (fig. 4)
- Per últim clica sobre *Translate sequence*

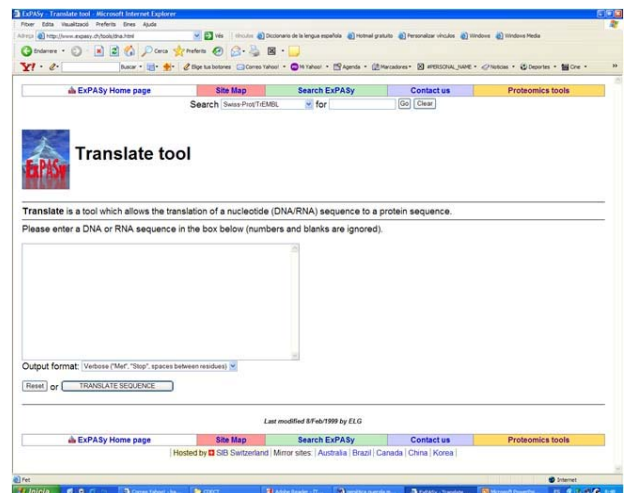


Figura 4. Pàgina amb l'eina *Translate Tool*

## Resultats

- c) Enganxa els resultats.

### Translate Tool - Results of translation

#### 5'3' Frame 1

SGSVQGVSPQGACRQASGLFIKLL  
 KLREDVDDEKNVNALAVCRSCCFH  
 IWCNACRGCANRRQFNDRCTEGH  
 DL **Stop** WTGENLCSSESLYIR **Stop** RFS  
 SGSEACIPAGSRGNFEKCYWCT  
 RSGWCALLRQL **Stop** YFQRC **Met** GRC  
 RRGCIDTEIIRNR **Stop** YYRRSSV  
**Stop** GLR **Stop** SVSDERFGYD **Stop** YQK  
 LPCG **Stop** YRQAGA AKWRNLLCG  
**Stop** HDTG **Stop** FQHIQREGFHYAHG  
 QQCISREDHEHTLLQSANAVQGLC  
 FRQDEPVRKYTVLTHACSKK **Stop** R  
 SG **Stop** SSSTTCSAVEENQIKRDEPV  
 N **Met** ASSRFDLVCLVTA **Met** NLYPYR  
 ITSLQAPQRTVCLG **Met** EQQNRYPR  
 SRICS **Stop** SHKRFPPACENI **Stop** RIP  
 AFFRLQ **Met** DGFLPRQRLVLP **Stop** **Stop**  
 RTGHQWFRLLVCSDWLAISTVHFRS  
 PSYNN **Stop** P

#### 5'3' Frame 2

(La part ombrejada en gris correspon a la proteïna buscada).

QAPYKG **Stop** ARRVPVDKPAACLSN  
 YLN **Stop** GR **Met** L **Met** **Met** KK **Met** LTLLLS  
 AGLVASIFGV **Met** PAAAAPT VVNSTI  
 VVPKGT TYDGQGKTFVANPSTLGD



GSQAENQKPVFRLEAGATLKNVIIG  
 APAADGVHCYGSCNISNVVWEDV  
 GEDALTLKSSGTVNITGGAAYKAY  
 DKVFQ **Met** NASGTINIKNFRADDIGK  
 LVRQNGGTSYAVN **Met** TLDNSNISN  
 VKDSI **Met** RTDSSVSQGGKITNTRYS  
 KVPTLFKGFASGKTSQSGNTQY  
**Stop** RTLAVKNKEAAEVQVPPAQP  
 RRTK **Stop** NG **Met** NQSTWPHVLIWF  
 VWLQP **Stop** ICIHIELLSRHLNEQFA  
 LGWSSRTAIQGAGFVHDLISGFLQ  
 HARIFEGYQRFSGYRWTGSCLGN  
 GFCLDKEQVTNGSGLFAQIGLPPFQ  
 QCISVLHRIITDQ

### [5'3' Frame 3](#)

RLRTRGEPAGCL **Stop** TSQRLVYQIT  
**Stop** IKGGC **Stop** **Stop** **Stop** KKC **Stop** RSC  
 CLPVLLLPYLV **Stop** CLPRLRQPSSI  
 QRSLYRRARP **Met** **Met** DRGKPL **Stop** R  
 IPLH **Stop** VTLKRRIRSLYSGWKQG  
 QL **Stop** K **Met** LLLVHPQR **Met** VCIVTAA  
 VIFPTLYGK **Met** **Stop** AR **Met** H **Stop** H  
**Stop** NHQEPLILPAEQRIRLTIKCFR  
**Stop** TLRVRLISKTSVR **Met** ISASWCG  
 K **Met** EEP **Met** RLT **Stop** HWIIPTYPT  
**Stop** RIPLCARTAVYLKGRSRTHVTP  
 KCQRCSRALLQARRASPEIHSINA  
 RLQ **Stop** KIKKRLKFKYHLLSR **Stop** G  
 EPNKTG **Stop** TSQHGLIPF **Stop** FGLF  
 GYSHESVSISNYFSPGTSTNSLPW  
 DGAAEPLSKEPDLF **Met** IS **Stop** AVSS  
 S **Met** REYLKDTSVFPATDGRVPASA  
 TGFALIKNRSP **Met** VPACLLRLACHF  
 NSAFPFSIV **Stop** **Stop** LTR

### [3'5' Frame 1](#)

SGQLLYDGERKCTVE **Met** ASQSEQT  
 SRNHW **Stop** PVLYQGKTRCRGRNPS  
 ICSRKNAGILQIFSHAGGNRL **Stop** D  
 HEQIRLLG **Stop** RFCCSIPRQTVR  
**Stop** GAWREVIRYGYRF **Met** AVTKQT  
 KSKRDEA **Met** LTGSSRFIWFSSAE  
 QVVLELQPLLYFLLQACVNTVYFRT  
 GSSCLKQSP **Stop** TALALWSNVCS  
**Stop** SSLEIHCCPCA **Stop** WNPSRWIC  
 WNPVSC **Stop** PHRRFLHFAAPACR  
 YHPHGSF **Stop** Y **Stop** SYPKRSSETLY  
 RKPYTLLRR **Stop** Y **Stop** RFL **Met** ISVS  
**Met** HPRHLPIQRWKYYSCRNNAHH  
 PLRVHQ **Stop** **Stop** HFSKLP LLPAGIQ  
 ASDSPLENRHL **Met** **Stop** RDSLQRF  
 PVHHRSCPSVQRSLN **Stop** RRLAQP  
 RQALHQIWKQQDRQTARALTFSS

STSSLNLSNLINKPLACLQAPCGLT  
 PCTEP

### [3'5' Frame 2](#)

LVSYYT **Met** ENGNALLKWQANLSKQ  
 AGTIGDLFFIKAKPVAEAGTRPSVA  
 GKTLSVSKYSR **Met** LEETAYEI **Met** N  
 KSGSLDSGSAAPSQGGKLFVEVPGE  
 K **Stop** FD **Met** DTDSWL **Stop** PNKPNQN  
 G **Met** RPC **Stop** LVHPVLF GSPQRLSR  
 WYLNFSRFFIFYCKRALILCISGLA  
 RLA **Stop** SKALEQRWHFGVTCVRDL  
 PLYTAVRAHNGILHVGYVGGIICH  
 VNRIGGSSILPHQLADIIRTEVFDIN  
 RTRSVHLKHFIVSLIRCSAGNINGS  
**Stop** **Stop** FQCQCILAYIFPYNVGNITA  
 AVT **Met** HTIRCGCTNNNIFQSCPCF  
 QPEYRLLILRLRTVT **Stop** CRGIRYK  
 GFPLSIIGRALRYNDR **Stop** IDDGWR  
 SRGRHYTKYGSNKTGRQQR **Stop** H  
 FFHHQHPPLI **Stop** VI **Stop** **Stop** TSRWL  
 VYRHPAGSPLVRS

### [3'5' Frame 3](#)

WSVIIRWRTE **Met** HC **Stop** NGKPI **Stop**  
 ANKPEPLVTCSLSRQNPLPRQEPV  
 HL **Stop** PEKRWYPSNILACWRKPL  
**Met** RS **Stop** TNPAPWIAVLLLHPKANC  
 SLRCLERSNSIWIQIHGCNQTNIK  
 TG **Stop** GHVDWFIPFYLVLLNG **Stop** A  
 GGT **Stop** TSAASLFFTASVR **Stop** YCV  
 FPDWLVLPEAKPLNSVGTLE **Stop** R  
 VFVIFP **Stop** DTLLSVRI **Met** ESFTLD  
**Met** LELSSV **Met** LTA **Stop** EVPPFCRTS  
 LPISSARKFLILIVPEAFI **Stop** NTLS  
**Stop** ALYAAPPVILTVPDDFSVNASS  
 PTSSHTTLEILQLP **Stop** QCTPSAAG  
 APIITFFKVAPASSRNTGF **Stop** FSA  
**Stop** EPSPNVEGFATKVFP CPS **Stop** V  
 VPFGT TIVELTTVGAAAAGITPN **Met**  
 EATRPADSKSVNIFFIINILP **Stop** FK  
**Stop** FDKQAAGLSTGTLRAHPLYGA  
**Stop**

- d) Quina és la seqüència d'aminoàcids que pot donar una proteïna, és a dir, que comenci amb un codó d'inici (ATG), acabi amb un codó STOP (TAA, TGA o TAT) i que codifiqui el polipeptid més llarg (Open Reading Frame, ORF)?

L'alumne ha d'enganxar la resposta correcta. La seqüència d'aminoàcids que comença per ATG, acaba en un codó STOP i codifica pel polipeptid més llarg és la que està senyalada en groc. La pro-

teïna podria començar per qualsevol de les tres metionines, però mirant la distància a seqüències consens, és més probable que comenci per la tercera.

e) En quina cadena es troba (5' → 3' o 3' → 5')?

La seqüència es troba en la cadena que va en direcció 5' → 3'

f) En quina pauta de lectura es troba (1a, 2a o 3a)?

La pauta de lectura és la segona.

### 3. Una vegada tenim la seqüència d'aminoàcids, hem de veure quina funció té.

Hem d'anar a una base de dades que compara la nostra seqüència d'aminoàcids amb altres seqüències de proteïnes ja conegudes. El resultat de la cerca de similituds és una llista ordenada de les proteïnes més semblants. Al costat de cadascuna apareix un número, *score*. Quant més alt és l'*score* (el número) més similitud hi ha.

A més obtindràs un gràfic de colors per a veure si l'homologia de seqüència és al llarg de tota la proteïna o només en un fragment localitzat, i els alineaments de la seqüència problema amb cada una de les seqüències de la llista anterior.

- Per arribar a la base de dades clica sobre l'adreça <http://www.expasy.ch/tools/>
- A continuació clica sobre *Similarity searches*
- A la nova pàgina clica sobre *BLAST at NCBI* (fig. 5)
- *Protein* (ja que tenim la seqüència d'aminoàcids)
- [Protein-protein BLAST \(blastp\)](#)
- Inserir la seqüència d'aminoàcids al requadre en blanc que trobaràs a la pàgina
- *Blast* (Obtenim les homologies)

g) Quin percentatge d'aminoàcids idèntics (*identities*) hi ha entre la teva seqüència problema i la més semblant? Per saber-ho pots clicar damunt del número *score*.

La resposta és 100%.

h) A quin organisme pertany la proteïna més semblant a la teva proteïna problema?

i) A la base de dades posa *Bacillus* BP-23. En el moment que es va seqüenciar el DNA encara

no s'havia classificat el bacteri. Ara se sap que és *Paenibacillus barcinonensis* (bacteri).



Figura 5. Pàgina per triar l'opció BLAST at NCBI

j) Quina és la funció hipotètica que té nostra proteïna?

La nostra seqüència d'aminoàcids té una elevada homologia amb la seqüència d'aminoàcids d'una pectat liasa (pectinasa). Per tant, la possible funció de nostra proteïna és la de trencar els enllaços que hi ha entre els monòmers del pectat o àcid poligalacturònic. Per saber si aquesta és la seva funció hauríem de fer assajos al laboratori.

k) Quin tipus d'enzim és? Què fa?

Com hem comentat abans, l'enzim és una pectat liasa. A la unitat d'enzims els alumnes han estudiat que és una liasa, per tant han de saber explicar què fa aquest enzim.

Les lises separen grups sense intervenció de l'aigua (sense hidròlisi), i generalment originen enllaços dobles a la molècula. Per tant, aquest enzim trenca els enllaços que hi ha entre els monòmers de l'àcid poligalacturònic o pectat formant un doble enllaç i sense alliberar aigua (fig. 6).

Relacionat amb aquesta pregunta es podria, comentar la diferència entre una liasa (pectat liasa) i una hidrolasa (poligalacturonasa).

Les hidrolases, també trenquen enllaços, però amb addició d'una molècula d'aigua que s'es-

cindeix i aporta un  $-OH$  a una part i un  $-H$  a l'altra, com es pot observar a la fig. 6.

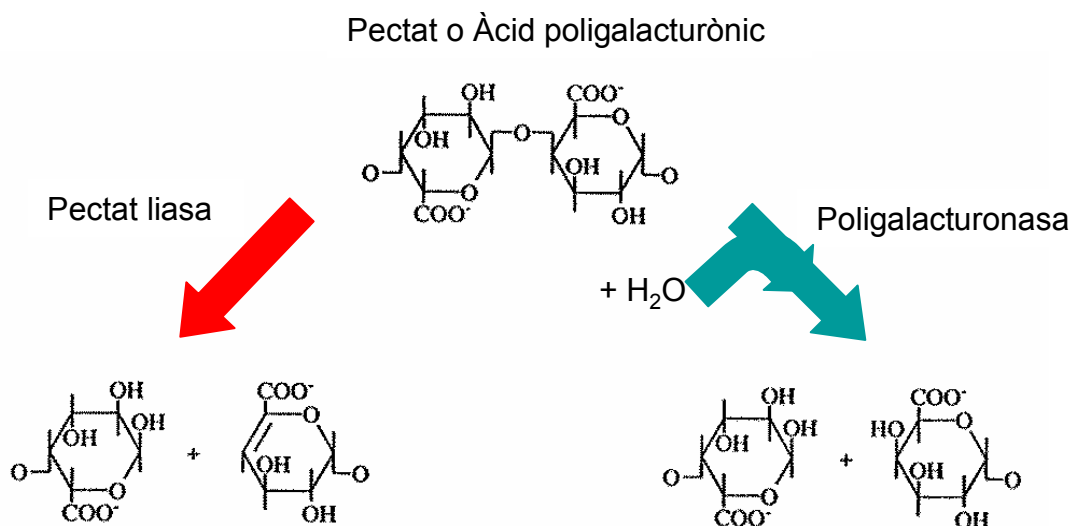
Felicitats, has aconseguit conèixer la funció d'una proteïna desconeguda!

### Redacció de la memòria

La redacció de la memòria és un treball personal. A continuació es proposen alguns suggeriments que poden resultar d'ajuda:

(S'aconsella al professor fer una posada en comú abans d'escriure la memòria).

- Ha de constar de introducció, material i mètodes, resultats, conclusió, discussió i bibliografia.
- Utilitzar el màxim rigor en les frases i expressions. El subjecte de l'oració ha de quedar sempre clar, evitant expressions que puguin donar lloc a més d'una interpretació.
- Poden copiar-se les preguntes dels exercicis en un document de Word, així, a l'hora de respondre un pot ajudar-se fent un copiar i pegar dels resultats obtinguts a través d'Internet.
- Assegurar la perfecta comprensió de les respostes.
- Afegir les adreces d'Internet utilitzades en la bibliografia.



**Figura 6.** Classificació de les pectinases que actuen sobre àcid poligalacturònic o pectat. Aquestes poden ser: Pectat liasa o Poligalacturonasa. L'enzim trobat trenca els enllaços que hi ha entre els monòmers de l'àcid poligalacturònic o pectat formant un doble enllaç i sense alliberar aigua.

### BIBLIOGRAFIA

- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. i Watson, J. *Biología molecular de la célula*. 3a ed. Barcelona: Omega, 1996.
- Griffiths, A. J. F., Gelbart, W. M., Miller, J. H. i Lewontin, R. C. *Genética Moderna*. Ed. McGraw-Hill Interamericana. 2000.
- Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C. i Gelbart, W. M.. *Genética*. Setena edició. Ed. McGraw-Hill Interamericana. 2002.
- Klug, W. W. i Cummings, M. R. *Conceptos de Genética*. Prentice Hall Iberia S. R. L. 1999.

Jimeno, A., Ballesteros, M. i Ugedo, L. *Biología. 2n Batxillerat*. Ed. Santillana. Barcelona, 2003.

Thain, M. i Hickman, M. *The penguin dictionary of biology*. Ed. The penguin books, 10a edició. 2000.

### GLOSSARI

**Exó:** Segments de RNA que contenen la informació per a la síntesi de proteïnes. Seqüències amb sentit.

**Gen:** Unitat fonamental, física i funcional, de la herència que transmet la informació d'una genera-



ció a la següent. Fragment d'àcid nucleic, generalment, ADN, que duu informació per a un caràcter.

**Genoma:** Material genètic complet d'una dotació cromosòmica.

**Intró:** Fragment del gen sense funció coneguda, que es transcriu però que el producte de la seva transcripció és eliminat del mRNA madur. Seqüència sense sentit que no codifica per cap proteïna.

**Organismes transgènics:** Organismes amb gens procedents d'altres organismes.

**Pauta de lectura** o *open reading frame* (ORF): És la porció del genoma d'un organisme que conté

una seqüència de bases que podria codificar potencialment una proteïna.

**Plasmidi:** Molècula extracromosòmica de DNA amb capacitat de replicació autònoma.

**SNPs** (Single Nucleotide Polymorphism): Variants polimòrfiques dels nucleòtids d'un gen.

**Splicing:** Procés d'eliminació d'introns i unió de les regions informatives o exons.

**Splicing alternatiu:** Procés d'eliminació d'introns i unió de les regions informatives o exons, que pot variar segons la proteïna que codifica.

## Annex: Pràctica per a l'alumnat

# Quina és la funció del gen clonat?

Margarita Soriano i Francesca Guerola

### Objectiu

En la següent pràctica volem replicar, transcriure i traduir un gen clonat en un plàsmid. D'aquesta manera podrem conèixer la seqüència d'aminoàcids i a partir d'ella identificar la seva funció.

A partir del fragment de DNA seqüenciat inserit en el plàsmid, hauràs de:

- replicar el fragment de DNA,
- transcriure la seqüència de nucleòtids,
- traduir la seqüència de nucleòtids,
- saber quina és la part del fragment de DNA que correspon al gen que volem investigar,
- i conèixer la funció de la proteïna.

Cal tenir en compte que el DNA bacterià no té introns. Això significa que cada gen codifica una proteïna.

### Materials

Les bases de dades i eines que necessitaràs per a resoldre l'activitat les pots trobar al full de pràctiques a *Protein Analysis System (ExPASy)*: <http://www.expasy.ch/tools/>. Aquesta serà la pàgina a la que pots dirigir-te per a triar l'eina adequada en cada ocasió i respondre les preguntes plantejades.

A més, has de tenir en compte que:

- Al llarg de la sessió has de guardar les adreces utilitzades per posar-les com a bibliografia.
- Cal guardar tota la informació que trobis en un disc o *pen drive* per a poder respondre així les preguntes i presentar un informe final de les pràctiques. (No has de deixar res teu en el disc dur de l'ordinador ja que l'endemà potser no li trobessis).

### Mètode

A partir d'una seqüència de nucleòtids hem de trobar el gen que volem investigar.

```
tcaggctccgtacaaggggtgagccccgaggggtcctgtagacaa
gccagcggctgttatcaaataacttaaaattaaggaggatggtgat
gatgaaaaaatgtaacgctctgctgctgccggctctgtgcttcc
atattggtgtaatgcctgccggtgagcgaaccgctgcaattcaa
cgatcgtgtaccgaagggcagacatgatggacaggggaaaa
cctttgtagcgaatccctctacattaggtgacggttctcaagcggagaa
tcagaagcctgtattccggctggaagcaggggcaacttgaaaaatg
ttattggtgcacccgcagcggatggtgtgattgacggcagctgt
aatattccaacgtgtatggaagatgtaggcgaggatgattgaca
ctgaaatcatcaggaaccgtaataattaccggcggagcagcgataa
ggcttacgataaagtgttcagatgaacgcttcgggtacgattaatc
aaaaactccgtgaggatgatcggcaagctggtgaggcaaaatg
gaggaacctcctatgaggtaacatgacactggataattccaacatc
ccaacgtgaaggattccattatgagcagcagcagcagctatctcaa
gggaagatcacgaacacacgttactccaagtgccaacgctgtca
agggcttgcctcaggcaagacgagccagtcgggaaatacacagta
ttaacgcagctgcagtaaaaaataaagaagcggctgaagtcaa
gtaccacctgctcagccgtgaggagaacaaataaaacgggatg
aaccagtcaacatggcctcatccgcttttgattggttgggttacagc
catgaatctgatccatcgaattactctcaccagcaccctcaacga
acagttgcttgggatggagcagcagaaccgctatccaaggagcc
ggattgttcagatctcataagcggtttccctcagcagcagcagcagc
gaaggataaccagcgttttccggctacagatggacgggttctgcctc
ggcaacgggtttgcctgataaagaacaggtcaccaatggtccgg
cttgttgcagattggctgcccattcaacagtcattccgttctccatc
gtataataactgaccaga
```

#### 1. Abans de començar la feina practicarem la replicació, transcripció i traducció a partir d'una petita seqüència de nucleòtids.

l) Copia la seqüència de nucleòtids assenyalada en negreta i fes la cadena complementària d'aquesta seqüència. A continuació transcriu i tradueix la seqüència utilitzant una pauta de lectura (fig. 1).

m) De la seqüència estudiada, on podria començar una hipotètica proteïna?

El DNA bacterià presenta una estructura circular de doble cadena. Això significa que el nostre gen pot estar en qualsevol de les dues cadenes. Com que les seqüències poden ser molt llargues, podem tenir la seqüència del gen en la cadena original o en la complementària, i a la vegada cadascuna pot tenir tres pautes de lectures o ORF (*open reading frame*), en els laboratoris es treballa amb programes que ens transcriuen i tradueixen la seqüència de la hipotètica proteïna.

n) A l'exercici a) has traduït la seqüència de mRNA. A continuació, torna a traduir la seqüència de mRNA utilitzant les altres 2 pautes de lectura. La seqüència d'aminoàcids és la mateixa?

## 2. La seqüència de nucleòtids completa (pàgina anterior) la replicarem, transcriurem i traduirem amb un programa informàtic per trobar la seqüència de la hipotètica proteïna:

- Primer, anirem a la següent pàgina: <http://www.expasy.ch/tools/>
- Clicarem sobre la següent opció: *DNA → Protein*
- A continuació farem dos clics sobre: *Translate*
- Enganxa la seqüència de nucleòtids en el quadre en blanc
- Per últim clica sobre *Translate sequence*

o) Enganxa els resultats

p) Quina és la seqüència d'aminoàcids que pot donar una proteïna, és a dir, que comenci amb un codó d'inici (ATG), acabi amb un codó de STOP (TAA, TGA o TAT) i que codifiqui el polipeptid més llarg (ORF)?

q) En quina cadena es troba (5' → 3' o 3' → 5')?

r) En quina pauta de lectura es troba (1a, 2a o 3a)?

## 3. Una vegada tenim la seqüència d'aminoàcids, hem de veure quina funció té

Hem d'anar a una base de dades que compara la nostra seqüència d'aminoàcids amb altres seqüències de proteïnes ja conegudes. El resultat de la cerca de similituds és una llista ordenada de les proteïnes més semblants. Al costat de cadascuna apareix un número, *score*. Com més gran és l'*score* (el número) més similitud hi ha.

A més obtindràs un gràfic de colors per veure si l'homologia de seqüència és al llarg de tota la proteïna o només en un fragment localitzat, i els alineaments de la seqüència problema amb cada una de les seqüències de la llista anterior.

- Per arribar a la base de dades clica sobre l'adreça <http://www.expasy.ch/tools/>
- A continuació clica sobre *Similarity searches*
- A la nova pàgina clica sobre *BLAST at NCBI*
- *Protein* (ja que tenim la seqüència d'aminoàcids)
- *Protein-protein BLAST (blastp)*
- Inserir la seqüència d'aminoàcids al requadre en blanc que trobaràs a la pàgina
- *Blast* (Obtenim les homologies)

s) Quin percentatge d'aminoàcids idèntics (*identities*) hi ha entre la teva seqüència problema i la més semblant? Per saber-ho pots clicar damunt del número *score*.

t) A quin organisme pertany la proteïna més semblant a la teva proteïna problema?

u) Quina és la funció hipotètica que té nostra proteïna?

v) Quin tipus d'enzim és? Què fa?

Felicitats, has aconseguit conèixer la funció d'una proteïna desconeguda!

## Redacció de la memòria

La redacció de la memòria és un treball personal. A continuació es proposen alguns suggeriments que poden resultar útils:

- Ha de constar d'introducció, material i mètodes, resultats, conclusió, discussió i bibliografia.
- Utilitzar el màxim rigor en les frases i expressions. El subjecte de l'oració ha de quedar sempre clar, evitant expressions que puguin donar lloc a més d'una interpretació.
- Poden copiar-se les preguntes dels exercicis en un document de Word. Així a l'hora de respondre hom pot fer servir el *copiar i enganxar* amb els resultats obtinguts a través d'Internet.
- Assegurar la perfecta comprensió de les respostes.
- Afegir les adreces d'Internet utilitzades en la bibliografia

		Segona lletra					
		U	C	A	G		
Primera lletra	U	UUU } Fen	UCU } Ser	UAU } Tir	UGU } Gln	U C A G	Tercera lletra
		UUC } Leu	UCC } Ser	UAC } Stop	UGC } Stop		
		UUA } Leu	UCA } Ser	UAA } Stop	UGA } Stop		
		UUG } Leu	UCG } Ser	UAG } Stop	UGG } Trp		
	C	CUU } Leu	CCU } Pro	CAU } His	CGU } Arg	U C A G	
		CUC } Leu	CCC } Pro	CAC } Gln	CGC } Arg		
		CUA } Leu	CCA } Pro	CAA } Gln	CGA } Arg		
		CUG } Leu	CCG } Pro	CAG } Gln	CGG } Arg		
	A	AUU } Ile	ACU } Tre	AAU } Asn	AGU } Ser	U C A G	
		AUC } Ile	ACC } Tre	AAC } Asn	AGC } Arg		
		AUA } Met	ACA } Tre	AAA } Lis	AGA } Arg		
		AUG } Met	ACG } Tre	AAG } Lis	AGG } Arg		
	G	GUU } Val	GCU } Ala	GAU } Asp	GGU } Gli	U C A G	
		GUC } Val	GCC } Ala	GAC } Asp	GGC } Gli		
		GUA } Val	GCA } Ala	GAA } Glu	GGA } Gli		
		GUG } Val	GCG } Ala	GAG } Glu	GGG } Gli		

**Figura 1.** Pauta per a la lectura de seqüències de mRNA.