Τμήμα Χημείας
Πανεπιστήμιο Κρήτης
Department of Chemistry
University of Crete

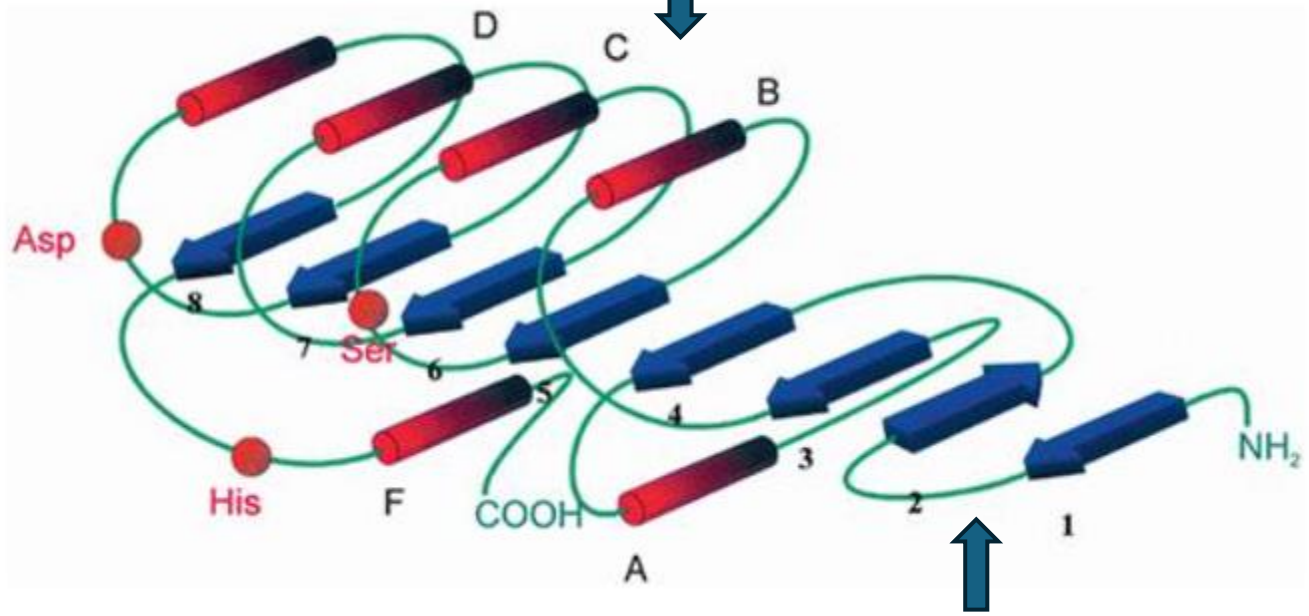
Εικονική κλωνοποίηση γονιδίου και δημιουργία βιβλιοθηκών μεταλλαγμάτων

Πρωτεϊνική Μηχανική - 2ο Σεμινάριο

Μανασσάκη Ελένη (ΑΜ:1301)
Καλφόγλου Παύλος Παναγιώτης (ΑΜ:1293)

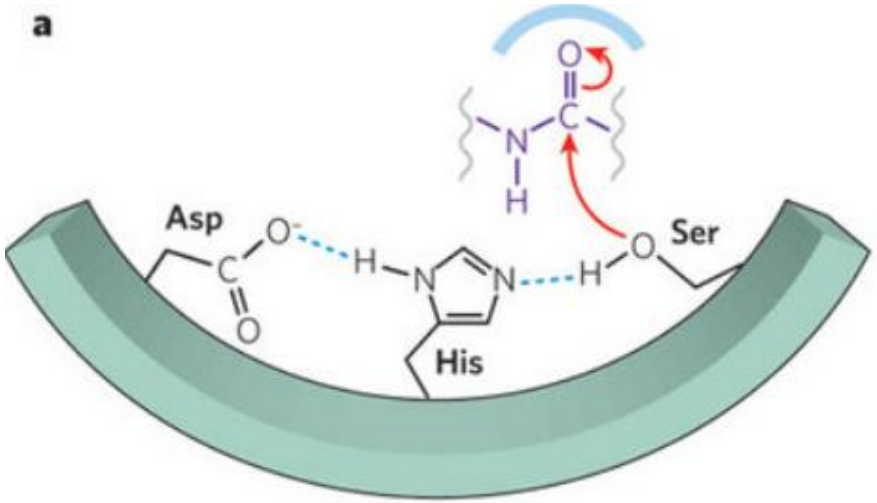
A/B υδρολάσες

α-έλικες



β-πτυχωτά φύλλα

Πυρηνόφιλη προσβολή



- Καταλυτική τριάδα Ser-His-Asp
- Συντηρημένο μοτίβο GXSXGE

- Bornscheuer, U.T. (2002). *Microbial carboxyl esterases: classification, properties and application in biocatalysis*. FEMS Microbiology Reviews, 26(1), 73–81. https://doi.org/10.1111/j.1574-6976.2002.tb00599.x
- Neitzel, J.J. (2010). *Enzyme Catalysis: The Serine Proteases*. Nature Education, 3(9):21.

Καρβοξυλο-εστεράση 2

Γενικές πληροφορίες	
Accession number (NCBI)	AAN66920.1 https://www.ncbi.nlm.nih.gov/protein/AAN66920.1
Entry name (UniProt)	Q88NB6_PSEPK https://www.uniprot.org/uniprotkb/Q88NB6/entry
Superfamily	Alpha-Beta Hydrolases
Subfamily (3DM)	1AUOB
Organism	Pseudomonas putida (KT2440)
Gene name	*estB*
Amino acids	218
Status	Predicted
Core identity	75%

Παράμετροι	
Soluprot	0.695
Theoretical pI	5.07
Molecular weight	24401.91
Ext. coefficient (cystines)	45045 1/M x cm
Extinction coefficient	44920 1/M x cm
Instability index	42.91

Καρβοξυλο-εστεράση 2

Αλληλουχία γονιδίου:

657bp

atgaccaacccgctgattctcgaaccacagaaaaccgctgacgcctgt
gtgatctggttgcacggtctgggggccgaccgttacgacttcctacct
gtggccgaattcatgcaggagcgcttgctcagcacacgcttcatcatg
ccccaggcaccgacccgccccgtgaccatcaacggcggctatgccatg
cccagctggtatgacatcaaggccatgaccccggcccgtgccatcgac
gaagcgcagctggaagaatctgccgagcaggtggttgccctgatcaag
gccgaacaggccaaaggcatcgacctgacgcggatcttcctggccggc
ttctcccagggcggcgcggtggtgctgcacactgcctatataaagtgg
caggaagcactgggcggggtgatcgccctgtccacctatgcccccacc
ttcaatgaccagcaccagttgagtgcctgccagcaacgcaccccggcc
ctgtgcctgcatggcgtgcatgactcggtggtcatcccgtccatgggc
cgtaccgccttcgagtacctgaacacctggggcgtcgctgcccgctgg
tacgaatacccgatggaacacgaagtggtggtcgaagaactgaacgat
atccacgactggctgagcaagcagctgcaatag

Αμινοξική αλληλουχία:

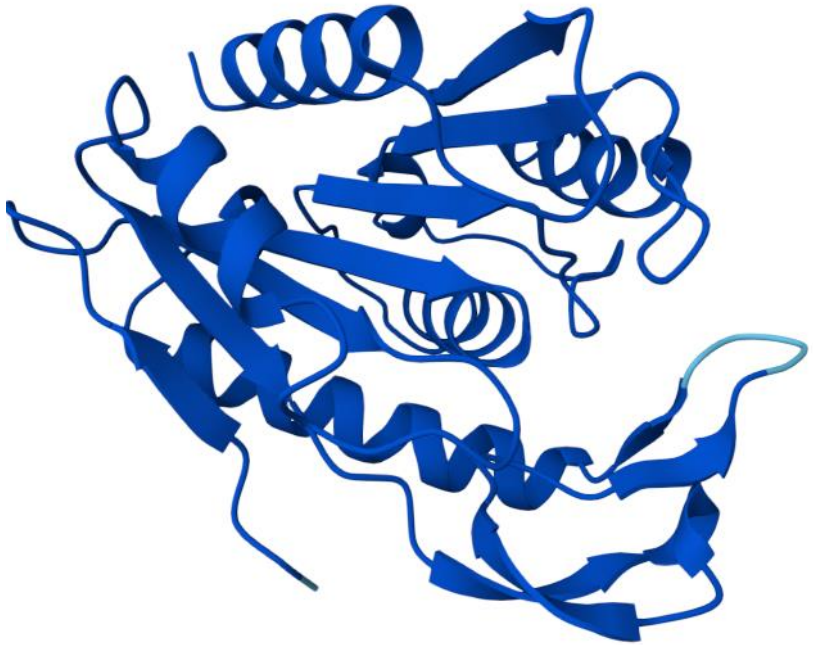
218aa

MTNPLILEPQKTADACVIWLHGLGADRYDF
LPVAEFMQERLLSTRFIMPQAPTRPVTING
GYAMPSWYDIKAMTPARAIDEAQLEESAEQ
VVALIKAEQAKGIDLTRIFLAGFSQGGAVV
LHTAYIKWQEALGGVIALSTYAPTFNDQHQ
LSACQQRTPALCLHGVHDSVVIPSMGRTAF
EYLNTWGVAARWYEYPMEHEVVVEELNDIH
DWLSKQLQ

- https://www.genome.jp/dbget-bin/www_bget?ppu:PP_1296
- https://www.uniprot.org/uniprotkb/Q88NB6/entry#sequences

Μοντέλα ομολογίας: AlphaFold

AlphaFold:
Model confidence: 96,94%
Πλήρης μοντελοποίηση (218 aa)



MolProbity evaluation

Clashscore: 0,88

Ramachandran:
outliers: 0,00%
favored: 98,15%

Poor rotamers: 0,56% (1)

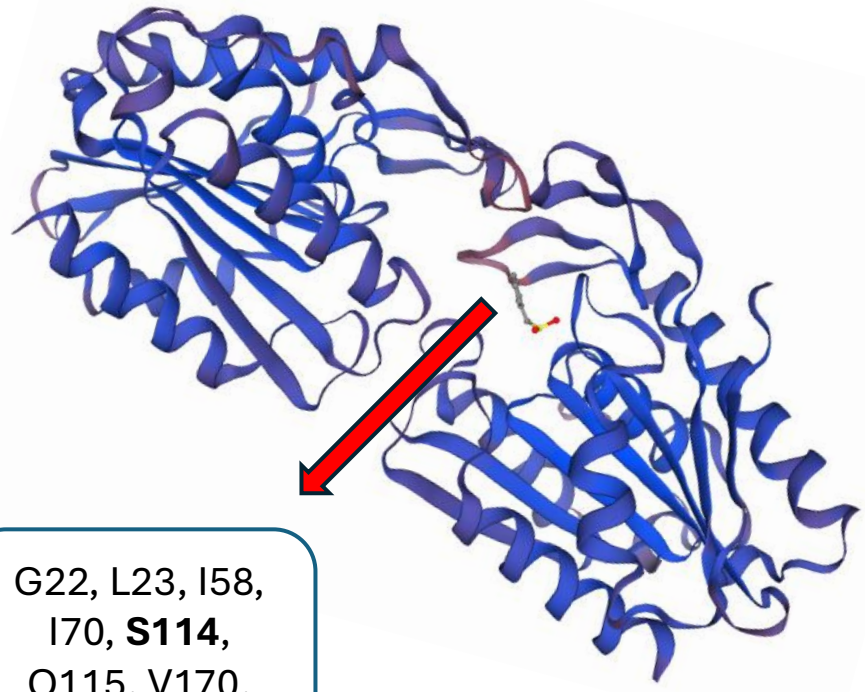
MolProbity score: 0,77

- Το μοντέλο παρουσιάζει εξαιρετική στερεοχημική ποιότητα, με χαμηλό clashscore και χωρίς Ramachandran outliers.

- https://alphafoldserver.com/fold/3ec7d9a14f36034d

Μοντέλα ομολογίας: Swiss model

Swiss model:
Πλήρης μοντελοποίηση 218 aa



G22, L23, I58, I70, **S114**, Q115, V170, V171, **H199**

MolProbity evaluation

Clashscore: 1,61

Ramachandran
outliers: 0,23% (Pro75)
favored: 95,37%

Poor rotamers: 2,78% (10)

MolProbity score: 1,57

- Το μοντέλο παρουσιάζει αποδεκτή γεωμετρία, αλλά η παρουσία Ramachandran και rotamers outliers υποδηλώνει χαμηλότερη ποιότητα σε σύγκριση με το AlphaFold μοντέλο.

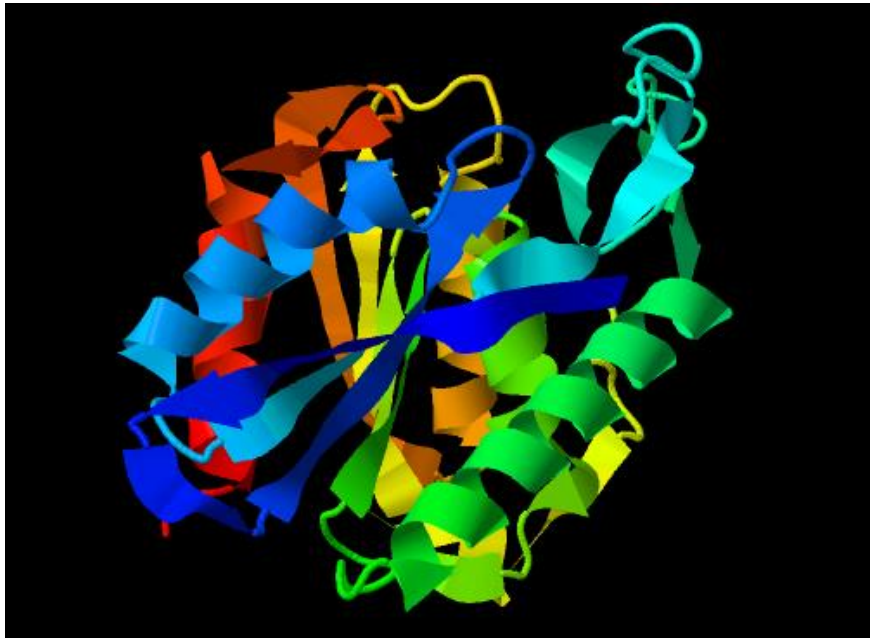
- https://swissmodel.expasy.org/interactive/TvJU6d/models/

Καταλυτικά αμινοξέα

- https://www.rcsb.org/sequence/1AUR

Μοντέλα ομολογίας: D-I-TASSER

D-I-TASSER:
Πλήρης μοντελοποίηση (218 aa)



Template	Predicted binding site residues
3fyuE	22,23,**114**,115,170,**199**
1auo0	58,59,62,64,72,73,74,75,167,**168**, 169,170,196,197,198,**199**,200
3f97A	22,23,5,30,113,**114**,115,**199**,200

- https://aideepmed.com/D-I-TASSER/output/D4807/

Ανάλυση 3DM

Καταλυτικό αμινοξύ	Αρίθμηση στην αλληλουχία	Αρίθμηση κατά 3DM
S	114	32
D	168	64
H	199	87

Αμινοξύ	Ποσοστά συντήρησης	
	% in Subfamily 1AUOB	% in Superfamily
S	100.00	84.15
D	99.78	75.13
H	99.78	97.71

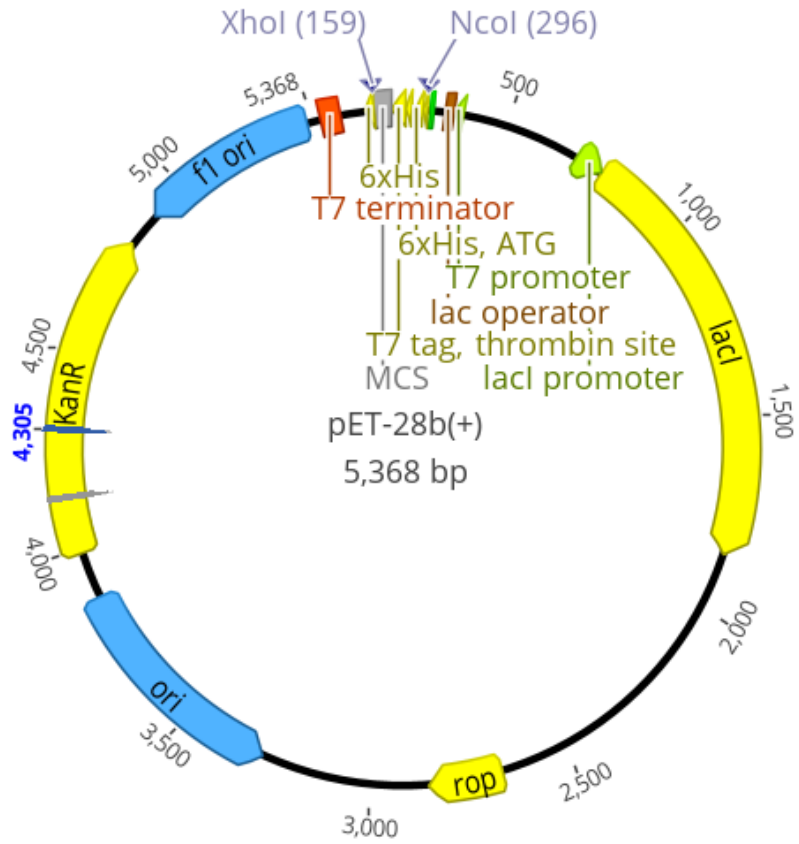
Αρίθμηση κατά 3DM	Αρίθμηση στην αλληλουχία	Πιο συχνό ζεύγος και ποσοστό	Ζεύγος στην αλληλουχία μας
27 - 48	109 - 134	G - A, 6.25%	F - G
53 - 64	139 - 168	A - D, 13.67%	S - D

Εικονική κλωνοποίηση

Βελτιστοποιημένη αλληλουχία:

```
atgacgaacccgctgattctggagccgcagaaaaccgccgatgcgtgc
gtgatttggctgcatggcctgggcgcggatcgctatgattttctgccg
gtggcggaatttatgcaggaacgcctgctgagcacccgtttcattatg
ccgcaggcgccgacccgcccggtgaccattaacggcggttacgccatg
ccgagctggtacgatatcaaagcgatgaccccggcgcgcgcgattgac
gaagcgcagctggaagaaagcgccgaacaggtggtggcgctgattaag
gccgaacaggcgaaaggcattgatctgacccgtatttttctggcgggc
ttcagccagggcggcgccgtggtactgcacaccgcgtatattaaatgg
caggaagcgctgggcggcgtgattgcgctgagcacctatgcgccgacc
ttcaacgatcagcatcagctgagcgcctgtcagcagcgcaccccggcg
ctgtgcctgcacggtgttcatgatagcgtggtgattccgagcatgggc
cgcaccgcgttcgaatatctgaacacctggggcgtggcggcgcgttgg
tatgaatacccgatggaacatgaagtggtagtggaagaactgaatgat
attcatgattggctgagcaaacagctgcagtaa
```

GC content: 61.49% → 58.75%
CAI : 0.42 → 0.95



NcoI: 5´CCATGG 3´

XhoI: 5´CTCGAG 3´

Εικονική κλωνοποίηση - Primers

Primers:

Forward primer 1:

5´GCG CCC CAT GG**T** **T**AT GAC GAA CC 3´

Forward primer 2:

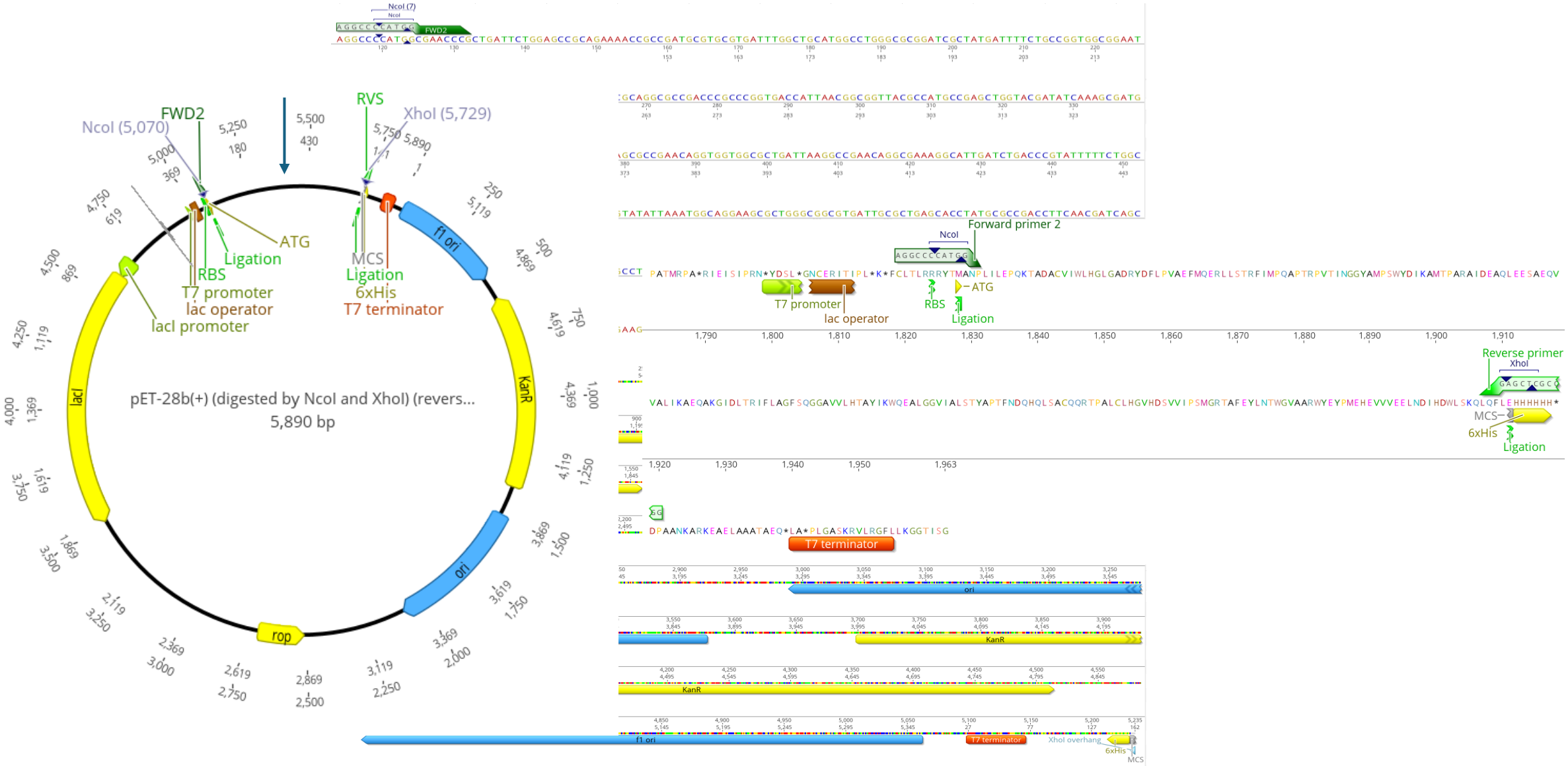
5´ AGG CCC CAT G**G**C GAA CCC G 3´

Reverse primer:

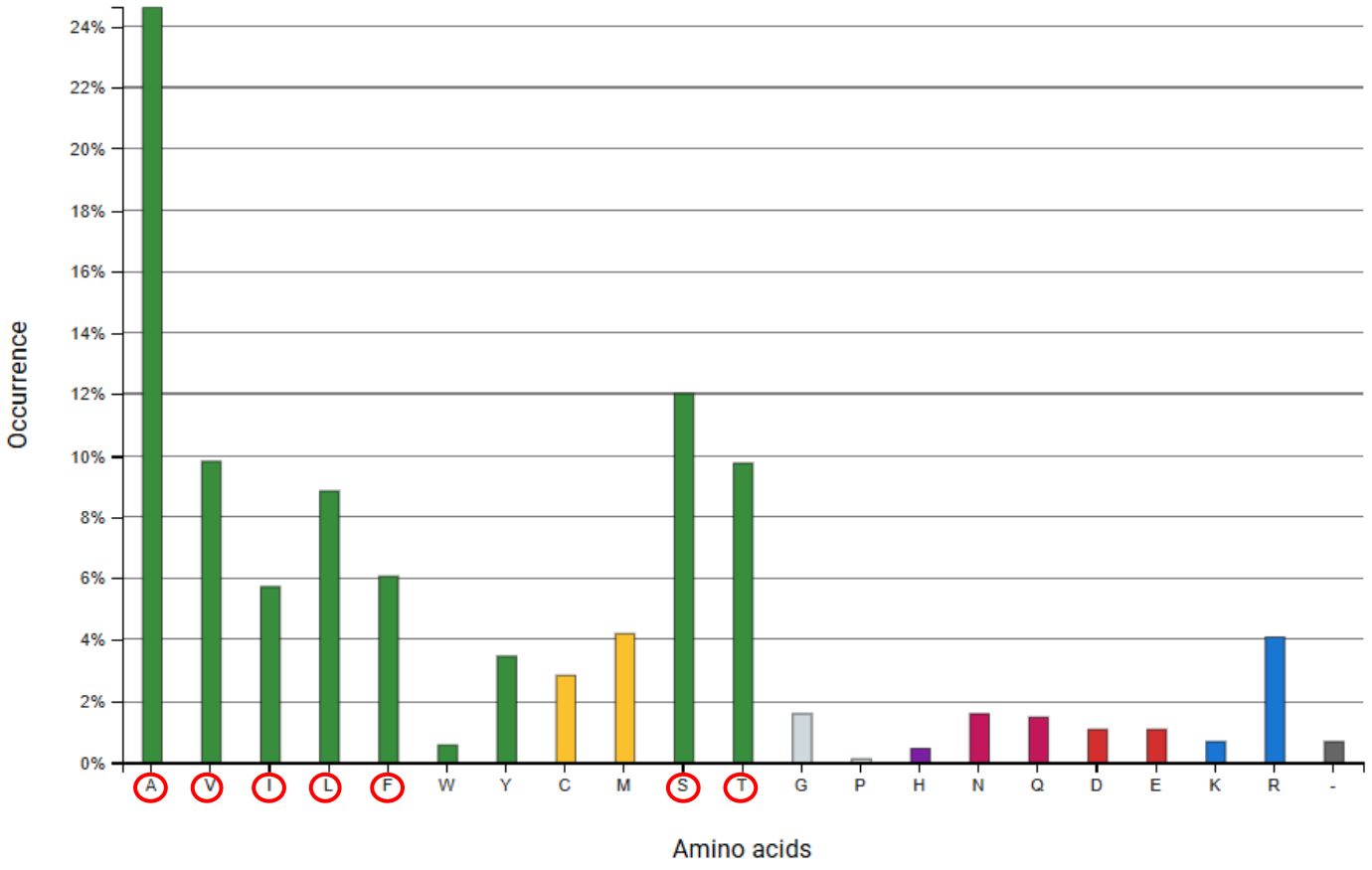
5´ GGC CGC TCG AG**A** **AA**C TGC AGC 3´

Primer	Length	GC (%)	T_M (°C)	T_M´(°C)
FW 1	23	61	61	68
RV	21	67	60	67
FW 2	19	74	60	66

Εικονική κλωνοποίηση



Βιβλιοθήκη – Θέση 22

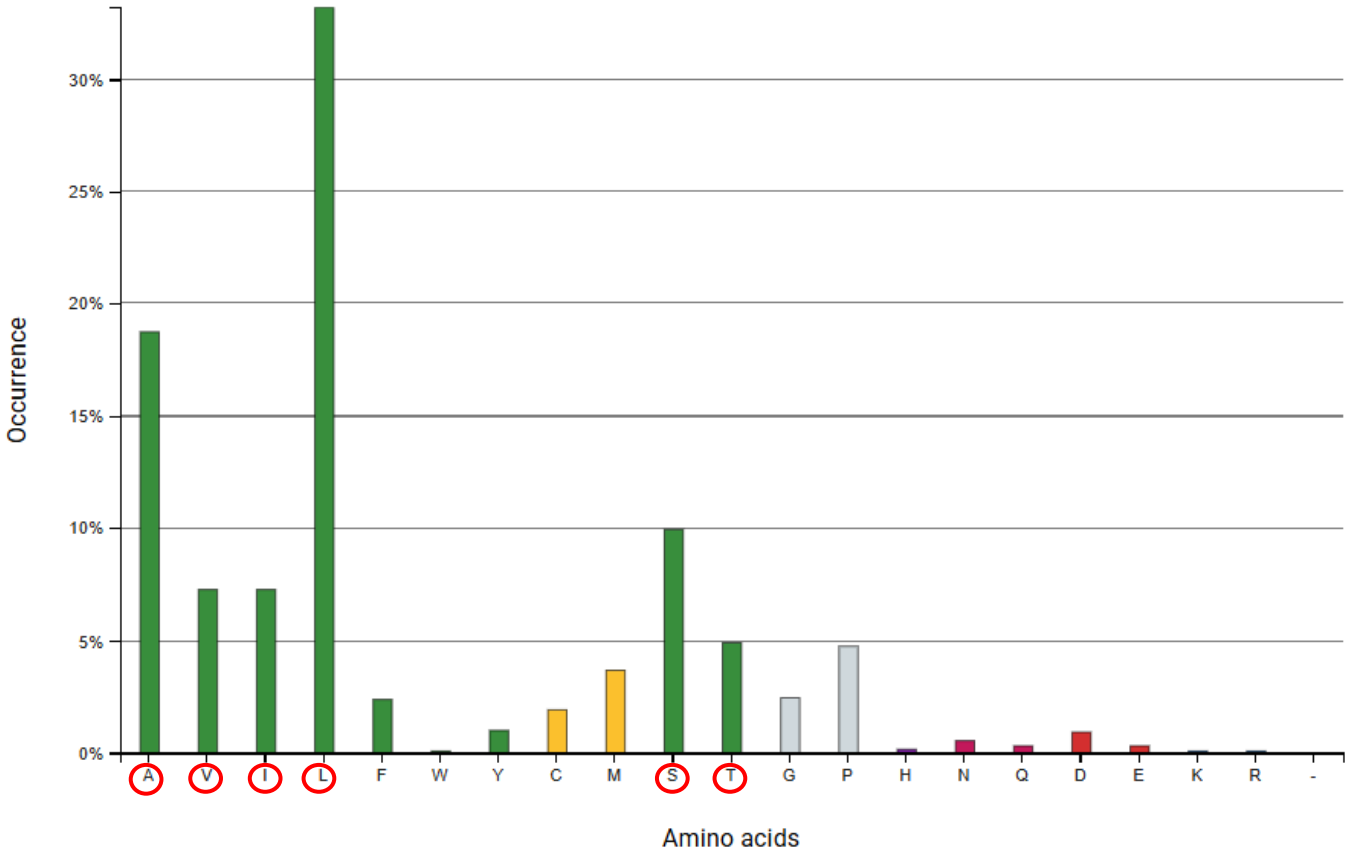


I. SELECT DEGENERATE CODON			
CODON	N	Y	T
DNA DEGENERACIES TABLE - IUBMB			
Base	Name	Bases	Comp
T	Thymine	T	A
C	Cytosine	C	G
A	Adenine	A	T
G	Guanine	G	C
Y	pYrimidine	C T	R
R	puRine	A G	Y
S	Strong (3H)	G C	S
W	Weak (2H)	A T	W
K	Keto	T G	M
M	aMino	A C	K
B	not A	C G T	V
D	not C	A G T	H
H	not G	A C T	D
V	not T	A C G	B
N	Unknown	A C G T	N

IV. Genetic Code - Codon Distribution

		SECOND POSITION OF THE CODON									
		T		C		A		G			
FIRST POSITION	T	TTT	Phe [F]	TCT	Ser [S]	TAT	Tyr [Y]	TGT	Cys [C]	T	THIRD POSITION
		TTC	Phe [F]	TCC	Ser [S]	TAC	Tyr [Y]	TGC	Cys [C]	C	
		TTA	Leu [L]	TCA	Ser [S]	TAA	Stp [end]	TGA	Stp [end]	A	
		TTG	Leu [L]	TCG	Ser [S]	TAG	Stp [end]	TGG	Trp [W]	G	
	C	CTT	Leu [L]	CCT	Pro [P]	CAT	His [H]	CGT	Arg [R]	T	
		CTC	Leu [L]	CCC	Pro [P]	CAC	His [H]	CGC	Arg [R]	C	
		CTA	Leu [L]	CCA	Pro [P]	CAA	Gln [Q]	CGA	Arg [R]	A	
		CTG	Leu [L]	CCG	Pro [P]	CAG	Gln [Q]	CGG	Arg [R]	G	
	A	ATT	Ile [I]	ACT	Thr [T]	AAT	Asn [N]	AGT	Ser [S]	T	
		ATC	Ile [I]	ACC	Thr [T]	AAC	Asn [N]	AGC	Ser [S]	C	
		ATA	Ile [I]	ACA	Thr [T]	AAA	Lys [K]	AGA	Arg [R]	A	
		ATG	Met [M]	ACG	Thr [T]	AAG	Lys [K]	AGG	Arg [R]	G	
	G	GTT	Val [V]	GCT	Ala [A]	GAT	Asp [D]	GGT	Gly [G]	T	
		GTC	Val [V]	GCC	Ala [A]	GAC	Asp [D]	GGC	Gly [G]	C	
		GTA	Val [V]	GCA	Ala [A]	GAA	Glu [E]	GGA	Gly [G]	A	
		GTG	Val [V]	GCG	Ala [A]	GAG	Glu [E]	GGG	Gly [G]	G	

Βιβλιοθήκη – Θέση 51

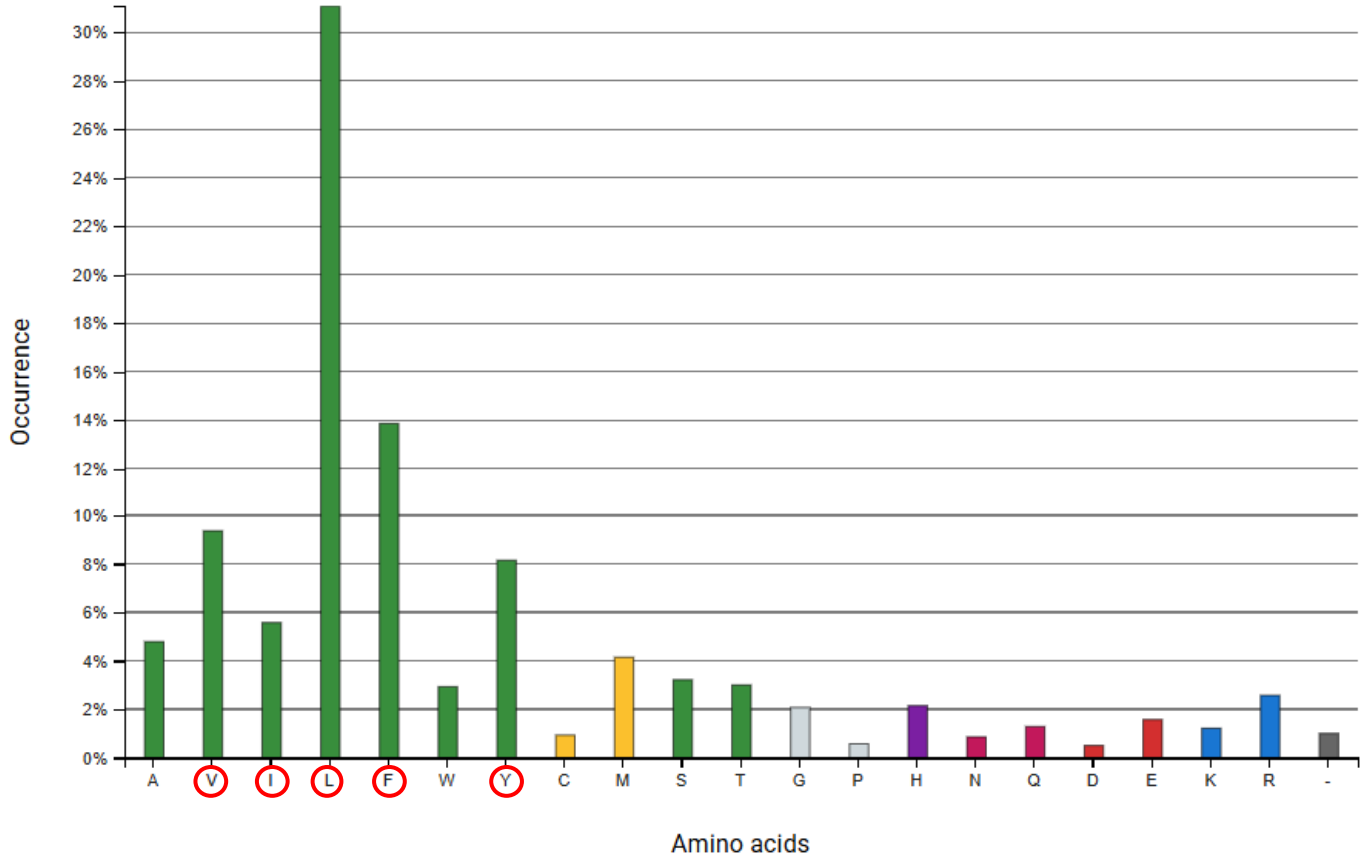


I. SELECT DEGENERATE CODON			
CODON	D	Y	A
DNA DEGENERACIES TABLE - IUBMB			
Base	Name	Bases	Comp
T	Thymine	T	A
C	Cytosine	C	G
A	Adenine	A	T
G	Guanine	G	C
Y	pYrimidine	C T	R
R	puRine	A G	Y
S	Strong (3H)	G C	S
W	Weak (2H)	A T	W
K	Keto	T G	M
M	aMino	A C	K
B	not A	C G T	V
D	not C	A G T	H
H	not G	A C T	D
V	not T	A C G	B
N	Unknown	A C G T	N

IV. Genetic Code - Codon Distribution

		SECOND POSITION OF THE CODON								
		T		C		A		G		
FIRST POSITION	T	TTT	Phe [F]	TCT	Ser [S]	TAT	Tyr [Y]	TGT	Cys [C]	T
		TTC	Phe [F]	TCC	Ser [S]	TAC	Tyr [Y]	TGC	Cys [C]	C
		TTA	Leu [L]	TCA	Ser [S]	TAA	Ter [end]	TGA	Ter [end]	A
		TTG	Leu [L]	TCG	Ser [S]	TAG	Ter [end]	TGG	Trp [W]	G
	C	CTT	Leu [L]	CCT	Pro [P]	CAT	His [H]	CGT	Arg [R]	T
		CTC	Leu [L]	CCC	Pro [P]	CAC	His [H]	CGC	Arg [R]	C
		CTA	Leu [L]	CCA	Pro [P]	CAA	Gln [Q]	CGA	Arg [R]	A
		CTG	Leu [L]	CCG	Pro [P]	CAG	Gln [Q]	CGG	Arg [R]	G
	A	ATT	Ile [I]	ACT	Thr [T]	AAT	Asn [N]	AGT	Ser [S]	T
		ATC	Ile [I]	ACC	Thr [T]	AAC	Asn [N]	AGC	Ser [S]	C
		ATA	Ile [I]	ACA	Thr [T]	AAA	Lys [K]	AGA	Arg [R]	A
		ATG	Met [M]	ACG	Thr [T]	AAG	Lys [K]	AGG	Arg [R]	G
	G	GTT	Val [V]	GCT	Ala [A]	GAT	Asp [D]	GGT	Gly [G]	T
		GTC	Val [V]	GCC	Ala [A]	GAC	Asp [D]	GGC	Gly [G]	C
		GTA	Val [V]	GCA	Ala [A]	GAA	Glu [E]	GGA	Gly [G]	A
		GTG	Val [V]	GCG	Ala [A]	GAG	Glu [E]	GGG	Gly [G]	G

Βιβλιοθήκη – Θέση 80



I. SELECT DEGENERATE CODON			
CODON	N	W	T
DNA DEGENERACIES TABLE - IUBMB			
Base	Name	Bases	Comp
T	Thymine	T	A
C	Cytosine	C	G
A	Adenine	A	T
G	Guanine	G	C
Y	pYrimidine	C T	R
R	puRine	A G	Y
S	Strong (3H)	G C	S
W	Weak (2H)	A T	W
K	Keto	T G	M
M	aMino	A C	K
B	not A	C G T	V
D	not C	A G T	H
H	not G	A C T	D
V	not T	A C G	B
N	Unknown	A C G T	N

IV. Genetic Code - Codon Distribution

FIRST POSITION	SECOND POSITION OF THE CODON: T		C		A		G		THIRD POSITION
T	TTT	Phe [F]	TCT	Ser [S]	TAT	Tyr [Y]	TGT	Cys [C]	T
	TTC	Phe [F]	TCC	Ser [S]	TAC	Tyr [Y]	TGC	Cys [C]	C
	TTA	Leu [L]	TCA	Ser [S]	TAA	Stop [end]	TGA	Stop [end]	A
	TTG	Leu [L]	TCG	Ser [S]	TAG	Stop [end]	TGG	Trp [W]	G
C	CTT	Leu [L]	CCT	Pro [P]	CAT	His [H]	CGT	Arg [R]	T
	CTC	Leu [L]	CCC	Pro [P]	CAC	His [H]	CGC	Arg [R]	C
	CTA	Leu [L]	CCA	Pro [P]	CAA	Gln [Q]	CGA	Arg [R]	A
	CTG	Leu [L]	CCG	Pro [P]	CAG	Gln [Q]	CGG	Arg [R]	G
A	ATT	Ile [I]	ACT	Thr [T]	AAT	Asn [N]	AGT	Ser [S]	T
	ATC	Ile [I]	ACC	Thr [T]	AAC	Asn [N]	AGC	Ser [S]	C
	ATA	Ile [I]	ACA	Thr [T]	AAA	Lys [K]	AGA	Arg [R]	A
	ATG	Met [M]	ACG	Thr [T]	AAG	Lys [K]	AGG	Arg [R]	G
G	GTT	Val [V]	GCT	Ala [A]	GAT	Asp [D]	GGT	Gly [G]	T
	GTC	Val [V]	GCC	Ala [A]	GAC	Asp [D]	GGC	Gly [G]	C
	GTA	Val [V]	GCA	Ala [A]	GAA	Glu [E]	GGA	Gly [G]	A
	GTG	Val [V]	GCG	Ala [A]	GAG	Glu [E]	GGG	Gly [G]	G

Βιβλιοθήκη

I. SIMULTANEOUS RANDOMIZATION OF DIFFERENT POSITIONS USING DIFFERENT DEGENERATE CODONS			
Position 1	N	Y	T
Position 2	D	Y	A
Position 3	N	W	T
Position 4			
Position 5			

II. SET % COVERAGE		95
and % WT background		0
Positions Randomized	Codons	Colonies
1		
1 + 2		
1 + 2 + 3	384	1149
1 + 2 + 3 + 4		
1 + 2 + 3 + 4 + 5		

IMPORTANT: When different positions are randomized simultaneously it is important to select in each position a degeneracy that contains the correspondent wild-type aminoacid, otherwise the evaluation of the possible effects of the individual libraries would not be properly analyzed, as one of the positions is forced to be mutated.

The creation of libraries randomizing more than one codon using **"mutation forced" degeneracies** is as well interesting as the diversity generated differs strongly to the wild-type amino acid sequence.

Amino acids [AA] encoded in each position					
	1	2	3	4	5
Ala [A]	1	1	0	0	0
Arg [R]	0	0	0	0	0
Asn [N]	0	0	1	0	0
Asp [D]	0	0	1	0	0
Cys [C]	0	0	0	0	0
Gln [Q]	0	0	0	0	0
Glu [E]	0	0	0	0	0
Gly [G]	0	0	0	0	0
His [H]	0	0	1	0	0
Ile [I]	1	1	1	0	0
Leu [L]	1	1	1	0	0
Lys [K]	0	0	0	0	0
Met [M]	0	0	0	0	0
Phe [F]	1	0	1	0	0
Pro [P]	1	0	0	0	0
Ser [S]	1	1	0	0	0
Thr [T]	1	1	0	0	0
Trp [W]	0	0	0	0	0
Tyr [Y]	0	0	1	0	0
Val [V]	1	1	1	0	0
Stop	0	0	0	0	0
Codons	8	6	8	0	0
AA	8	6	8	0	0

Βιβλιοθήκες NNK & NDT

I. SELECT DEGENERATE CODON

CODON	N	N	K

DNA DEGENERACIES TABLE - IUBMB

Base	Name	Bases	Comp
T	Thymine	T	A
C	Cytosine	C	G
A	Adenine	A	T
G	Guanine	G	C
Y	pYrimidine	C T	R
R	puRine	A G	Y
S	Strong (3H)	G C	S
W	Weak (2H)	A T	W
K	Keto	T G	M
M	aMino	A C	K
B	not A	C G T	V
D	not C	A G T	H
H	not G	A C T	D
V	not T	A C G	B
N	Unknown	A C G T	N

II. SET % COVERAGE	95
and % WT background	0

AA Sites	Codons*	Colonies**
1	32	94
2	1024	3066
3	32768	98163
4	1048576	3141251
5	33554432	100520093
10	1,13E+15	3,4E+15

References

Codons included in the selected degenerated codon (and amino acids encoded)

III. Amino acids [AA]

	Codons
Ala [A]	2
Arg [R]	3
Asn [N]	1
Asp [D]	1
Cys [C]	1
Gln [Q]	1
Glu [E]	1
Gly [G]	2
His [H]	1
Ile [I]	1
Leu [L]	3
Lys [K]	1
Met [M]	1
Phe [F]	1
Pro [P]	2
Ser [S]	3
Thr [T]	2
Trp [W]	1
Tyr [Y]	1
Val [V]	2
Stop	1
Codons	32
AA	20

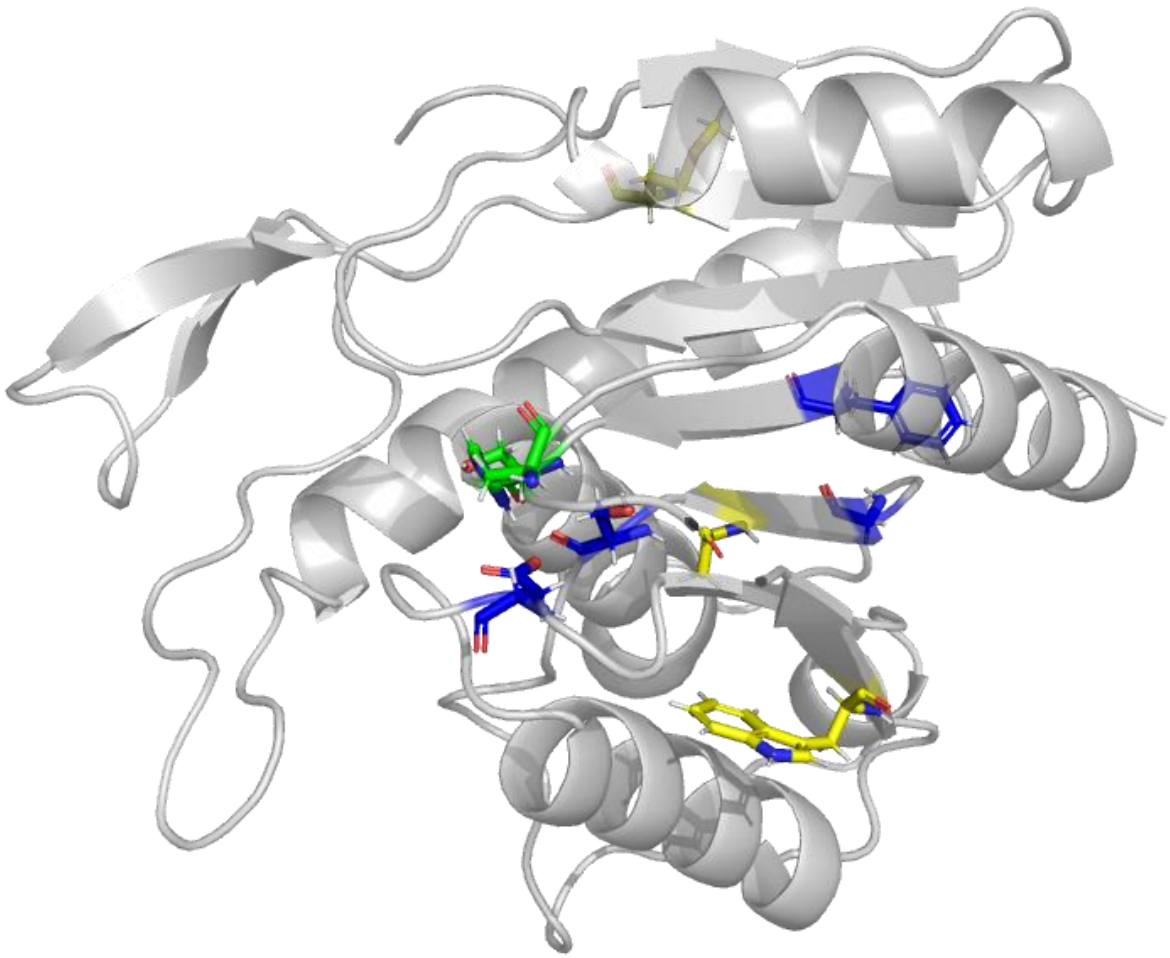
IV. Genetic Code - Codon Distribution

FIRST POSITION	SECOND POSITION	THIRD POSITION	T		C		A		G	
T	T	T	TTT	Phe [F]	TCT	Ser [S]	TAT	Tyr [Y]	TGT	Cys [C]
		C	TTC	Phe [F]	TCC	Ser [S]	TAC	Tyr [Y]	TGC	Cys [C]
		A	TTA	Leu [L]	TCA	Ser [S]	TAA	*Ter* [end]	TGA	*Ter* [end]
		G	TTG	Leu [L]	TCG	Ser [S]	TAG	*Ter* [end]	TGG	Trp [W]
	C	T	CTT	Leu [L]	CCT	Pro [P]	CAT	His [H]	CGT	Arg [R]
		C	CTC	Leu [L]	CCC	Pro [P]	CAC	His [H]	CGC	Arg [R]
		A	CTA	Leu [L]	CCA	Pro [P]	CAA	Gln [Q]	CGA	Arg [R]
		G	CTG	Leu [L]	CCG	Pro [P]	CAG	Gln [Q]	CGG	Arg [R]
	A	T	ATT	Ile [I]	ACT	Thr [T]	AAT	Asn [N]	AGT	Ser [S]
		C	ATC	Ile [I]	ACC	Thr [T]	AAC	Asn [N]	AGC	Ser [S]
		A	ATA	Ile [I]	ACA	Thr [T]	AAA	Lys [K]	AGA	Arg [R]
		G	ATG	Met [M]	ACG	Thr [T]	AAG	Lys [K]	AGG	Arg [R]
	G	T	GTT	Val [V]	GCT	Ala [A]	GAT	Asp [D]	GGT	Gly [G]
		C	GTC	Val [V]	GCC	Ala [A]	GAC	Asp [D]	GGC	Gly [G]
		A	GTA	Val [V]	GCA	Ala [A]	GAA	Glu [E]	GGA	Gly [G]
		G	GTG	Val [V]	GCG	Ala [A]	GAG	Glu [E]	GGG	Gly [G]

V. Amino acid Distribution - Chemical Classification

Side Chain	Amino acids	%
Acidic [-]	Asp, Glu	6,3
Basic [+]	Arg, His, Lys	15,6
Non-polar aliphatic	Ala, Ile, Leu, Met, Val	28,1
Aromatic	Phe, Trp, Tyr	9,4
Polar	Asn, Cys, Gln, Ser, Thr	25,0
Special Features	Gly, Pro	12,5

Μοντέλο



Βιβλιογραφία

- Di Cera, E. (2009). Serine proteases. *IUBMB Life*, 61(5), 510–515. https://doi.org/10.1002/iub.186
- Verma, S., Dixit, R. & Pandey, K.C. (2016). Cysteine proteases: modes of activation and future prospects as pharmacological targets. *Frontiers in Pharmacology*, 7, 107. https://doi.org/10.3389/fphar.2016.00107
- Davies, G. & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, 3(9), 853–859. https://doi.org/10.1016/S0969-2126(01)00220-9
- Bornscheuer, U.T. (2002). *Microbial carboxyl esterases: classification, properties and application in biocatalysis*. FEMS Microbiology Reviews, 26(1), 73–81. https://doi.org/10.1111/j.1574-6976.2002.tb00599.x
- Neitzel, J.J. (2010). *Enzyme Catalysis: The Serine Proteases*. Nature Education, 3(9):21.

Σύνδεσμοι

- https://www.ncbi.nlm.nih.gov/protein/AAN66920.1
- https://www.genome.jp/dbget-bin/www_bget?ppu:PP_1296
- https://swissmodel.expasy.org/interactive/TvJU6d/models/
- https://alphafoldserver.com/fold/3ec7d9a14f36034d
- https://www.rcsb.org/sequence/1AUR
- https://www.uniprot.org/uniprotkb/Q88NB6/entry#sequences
- https://en.vectorbuilder.com/tool/codon-optimization/4ec9a942-2353-4275-a815-7233572f9aa7.html
- https://www.biosyn.com/Gizmo/Tools/Oligo/Oligonucleotide%20Properties%20Calculator.htm
- http://nebcloner.neb.com/#!/redigest

Ευχαριστούμε για την προσοχή σας!